Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 5708746, 15 pages
https://doi.org/10.1155/2018/5708746

Research Article

Robust Semiautomatic 2D-to-3D Conversion with Welsch M-Estimator for Data Fidelity

Hongxing Yuan, Shaoqun Wu, Peng An, Chunya Tong, You Zheng, Shudi Bao, and Yongping Zhang

Ningbo University of Technology, Ningbo, China

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

Received 26 July 2017; Revised 12 February 2018; Accepted 27 February 2018; Published 4 April 2018

Academic Editor: Simone Bianco

Semiautomatic 2D-to-3D conversion plays an important role in generating 3D contents for display. However, most existing methods assume that user scribbles are perfectly correct, and only give acceptable results when user provides accurate labels. To address this problem, Welsch M-estimator data fidelity is used to resist erroneous scribbles. The Welsch M-estimator data fidelity which is able to alleviate the influence of inaccurate scribbles has theoretical guarantee by means of its redescending property. First, the Welsch M-estimator is introduced to measure the fidelity between estimated depth and user provided depth; then local smoothness is built by using color weighted Welsch M-estimator to make neighboring pixels with similar colors have similar depth values. Finally, we solve the problem using generalized iteratively reweighted least squares algorithm. Experiments demonstrate that our method obtains competitive performance in the absence of inaccurate scribbles and outperforms the state of the art both visually and quantitatively in the presence of inaccurate scribbles.

1. Introduction

3D videos have gained much attention as 3D viewing became popular and Virtual Reality (VR) market emerged. The biggest issue of 3D industry is lack of program material. 2D-to-3D conversion is a practical solution to alleviate such content shortage by estimating depth information from monoscopic images [1]. High quality depth extraction plays a key role in 2D-to-3D conversion [2].

Depending on whether human intervention is utilized, 2D-to-3D conversion can be divided into three categories: manual, automatic, and semiautomatic method [3]. Manual method can provide high quality results with per-pixel depth assignment by labeling; thus, this makes the process of conversion both cumbersome and expensive [4]. Automatic method attempts to estimate depth from monoscopic images utilizing various cues such as defocus, texture gradients, and scattering [5]. Recently, deep-learning-inspired approaches have been proposed for automatically converting 2D video/image to 3D format [5–10]. Although these methods can produce depth maps automatically, they are hard to provide robust and stable conversion results in any general content. Semiautomatic method can balance 3D quality with conversion cost which consists of the following steps: first let user label on chosen key frames to provide sparse depth, then obtain dense depth maps of key frames via sparse-to-dense propagation, and finally generate depth maps of nonkey frames by depth propagation from the key frames [3]. The conversion quality largely depends on the accuracy of depth maps for key frames. Therefore, we focus on the most relevant work of semiautomatic 2D-to-3D conversion about sparse-to-dense depth propagation for key frames.

Various methods have been proposed for dense depth estimation from user scribbles. Rzeszutek et al. [11] exploit random walks (RW) to generate dense depth based on the user input, but RW has problems in preserving strong edges [12], thus resulting in blurring artifacts at object boundaries. Phan and Androutsos [12] attempt to enhance depth discontinuities of RW by introducing the hard segmentation constraints provided by graph-cuts (GC). However, GC is hard to locate object boundaries at the transition

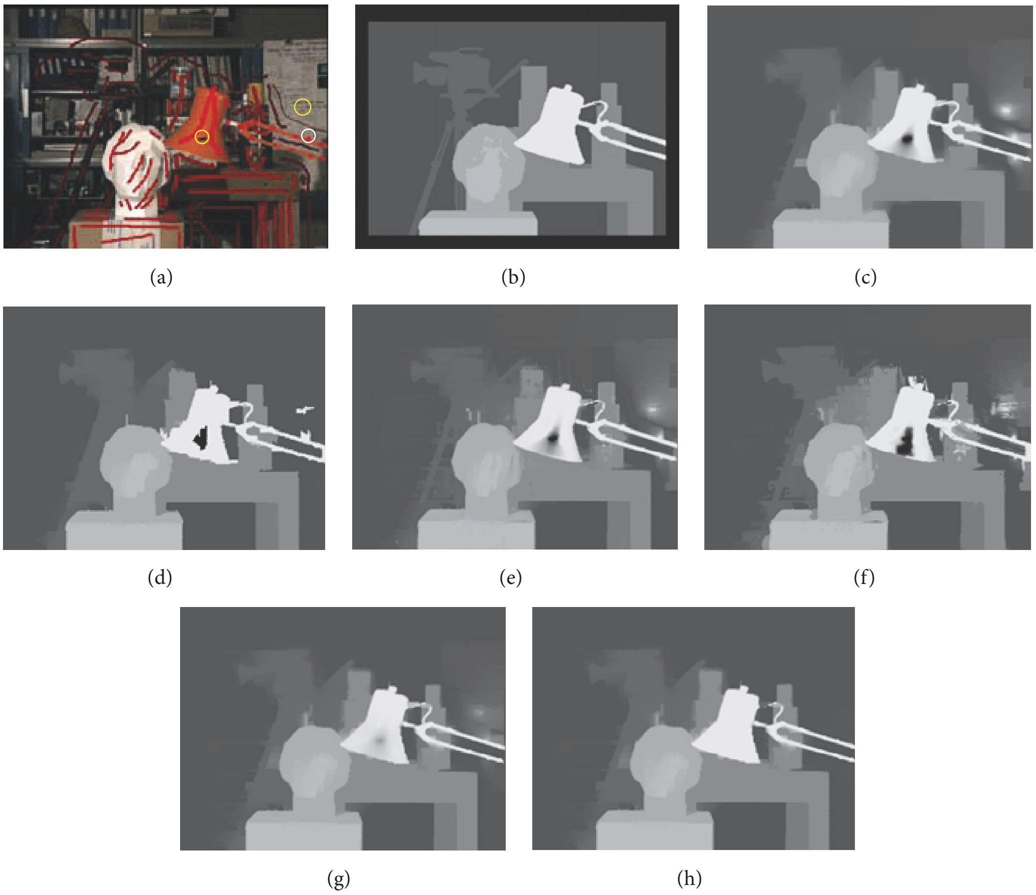


Figure 1: Depth estimation with erroneous user input, where (a) is user labeled image (erroneous scribbles at object boundaries are marked by white circles, and errors inside objects or background are marked by yellow circles), (b) is ground truth depth, (c) is result of Rzeszutek et al. [11], (d) is result of Phan and Androutsos [12], (e) is result of Yuan et al. [14], (f) is result of Wu et al. [15], (g) is result of Ham et al. [24], and (h) is result of the proposed method.

from foreground to background with low contrast [13] and may introduce fake boundaries. Our previous work [14] demonstrates that depth discontinuities of RW can be enhanced with nonlocal pairwise constraints. Wu et al. [15] enhance depth boundaries with superpixel constraints which can prevent depth propagation across low contrast edge regions. Lopez et al. [16] formulate depth estimation from user scribbles as a graph based optimization problem with equality, inequality, and perspective constraints. Becker et al. [17] let user annotate depth discontinuities in key frames and learn depth edges of nonkey frames with random forests, which can produce dense depth maps with sharp edges at discontinuities but with more cumbersome labeling work. Kawai and Sasaki [18] propose to generate dense depth from user provided anchor points on the outline of objects at key frames, but this increases user labeling difficulties since it is hard to locate the outline of objects. Donatsch et al. [19] employ user provided geometric features to generate stereo pairs directly, but mainly suitable for images with buildings. Zhang et al. [20] utilize interactive segmentation to refine foreground depth, but inaccurate segmentation may introduce depth artifacts. Iizuka et al. [21] show that geodesic distance based interpolation can obtain dense depth efficiently from user input with few scribbles. Liao et al. [22] let user assign diffusion strength during sparse-to-dense propagation to influence the depth estimation.

Existing approaches mainly focus on enhancing depth quality and assume that user scribbles are entirely accurate. Therefore, they generate correct depth only with accurate user scribbles, and even small errors in the input may degrade the depth quality significantly as shown in Figure 1. The erroneous scribbles inside objects or background can be easily removed by users during conversion process. However, it is hard for users to make adjustments when erroneous input appears at object boundaries. The user friendly semiautomatic 2D-to-3D conversion method should have the ability to remove erroneous input automatically. Handling of inaccurate user labels has been addressed in semiautomatic image segmentation [23, 25, 26]. While Subr et al. [25] and Bai and Wu [26] can discriminate accurate and inaccurate input, they focus on binary labels which cannot be applied to 2D-to-3D conversion directly. Oh et al. [23] utilize occurrence and cooccurrence probability (OCP) of color values for labeled pixels to estimate the reliability of each label, but may mistake correct labels

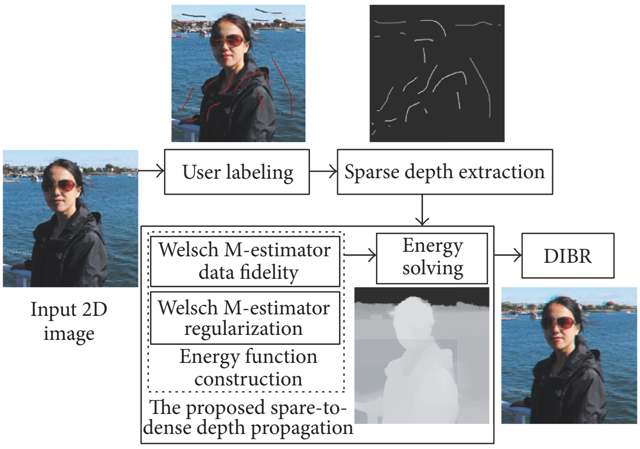

Input 2D image
User labeling
Sparse depth extraction
Welsch M-estimator data fidelity
Welsch M-estimator regularization
Energy function construction
The proposed spare-to-dense depth propagation
Energy solving
DIBR

Figure 2: Semiautomatic 2D-to-3D conversion based on the proposed method.

for incorrect labels. Surprisingly, there are few 2D-to-3D conversion methods to handle inaccurate user labels. To address this issue, we propose a robust method based on Welsch M-estimator for data fidelity motivated by the fact that Welsch loss based redescending M-estimator can be efficiently resistant to extreme outliers [27]. We note that Welsch M-estimator has been used to construct the regularizer for depth superresolution in recent years [24, 28–31]. Although employing Welsch M-estimator for regularization handles structural difference between texture and depth images like these methods, we leverage it for data fidelity to resist the influence of inaccurate input on the estimated depth.

Thanks to Welsch M-estimator for data fidelity, our approach outperforms existing methods in the presence of inaccurate input and provides at least comparable performance in the absence of erroneous input. The remainder of this paper is divided into three sections. In Section 2, our method for robust semiautomatic 2D-to-3D conversion is presented. Experimental results are provided in Section 3. Finally, we give conclusion in Section 4.

2. Proposed Approach

The semiautomatic 2D-to-3D conversion framework based on the proposed method is shown in Figure 2. First, we provide an interaction tool (https://github.com/tcyhx/brush2depth) for user to brush sparse scribbles on input 2D images or key frames, indicating initial depth. Second, sparse depth map is obtained from the intensities of user scribbles, where lighter and darker denote closer and farther from the viewer, respectively. Third, we construct data fidelity and regularization terms using Welsch M-estimator and formulate the sparse-to-dense depth propagation as a robust optimization, which will be illustrated in Section 2.1. Then, we solve the optimization problem via generalized iteratively reweighted least squares (IRLS) [32], which will be discussed in Section 2.2. Finally, we produce 3D content using a depth image based rendering (DIBR) technique proposed in our previous work [33].

2.1. Model. Let Ω denote a set containing user labeled pixel locations. Given the n-pixel input image I, user provided sparse depth map u, and the estimated dense depth map d, we denote by I_i, u_i, and d_i the corresponding values of pixel i. Without loss of generality, we assume that I_i, u_i, and d_i are normalized in the range 0 to 1. We minimize the following objective function to estimate d from u:

$$\varepsilon(d) = \underbrace{\sum_{i\in\Omega}\varphi_{\eta_1}\left(d_i - u_i\right)}_{\text{data fidelity term}} + \lambda\underbrace{\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_i} w_{ij}\varphi_{\eta_2}\left(d_i - d_j\right)}_{\text{regularization term}}, \quad (1)$$

where $\varphi_\eta(x) = (1/\eta)(1 - e^{-\eta x^2})$ denotes Welsch function and η is a bandwidth parameter which has influence on the strength of penalty to outliers, $\mathcal{N}_i$ is the local neighboring index set of the pixel i, w_{ij} represents Gaussian weighting function measuring appearance similarities between pixel i and j, which is given by $w_{ij} = e^{-\mu(I_i - I_j)^2}$, and λ is the parameter to balance data fidelity with regularizer.

It can be seen from formula (1) that we introduce the data consistency by Welsch's function to suppress user erroneous input while adapting the Welsch loss for regularization. Since Welsch M-estimator can deal with outliers with large magnitudes [27], we can ignore inaccurate scribbles with the data fidelity term while minimizing depth blurring caused by structural differences between texture and depth images via the regularizer. Recently, Ham et al. [24], Kim et al. [28, 29], and Liu et al. [30, 31] have introduced the regularity of depth maps by Welsch M-estimator. Our model differs from [24, 28–31] in its data fidelity. These methods all use a quadratic data fidelity, which cannot handle inaccuracies in user scribbles. As shown in Figure 3, Welsch M-estimator data fidelity can help to reduce visual artifacts caused by inaccurate input but quadratic data fidelity cannot suppress erroneous input. The characteristics of our model will be further illustrated in Section 2.3.

2.2. Solver. The optimization problem to minimize (1) is nonconvex, and can be solved by the GIRLS algorithm [32].

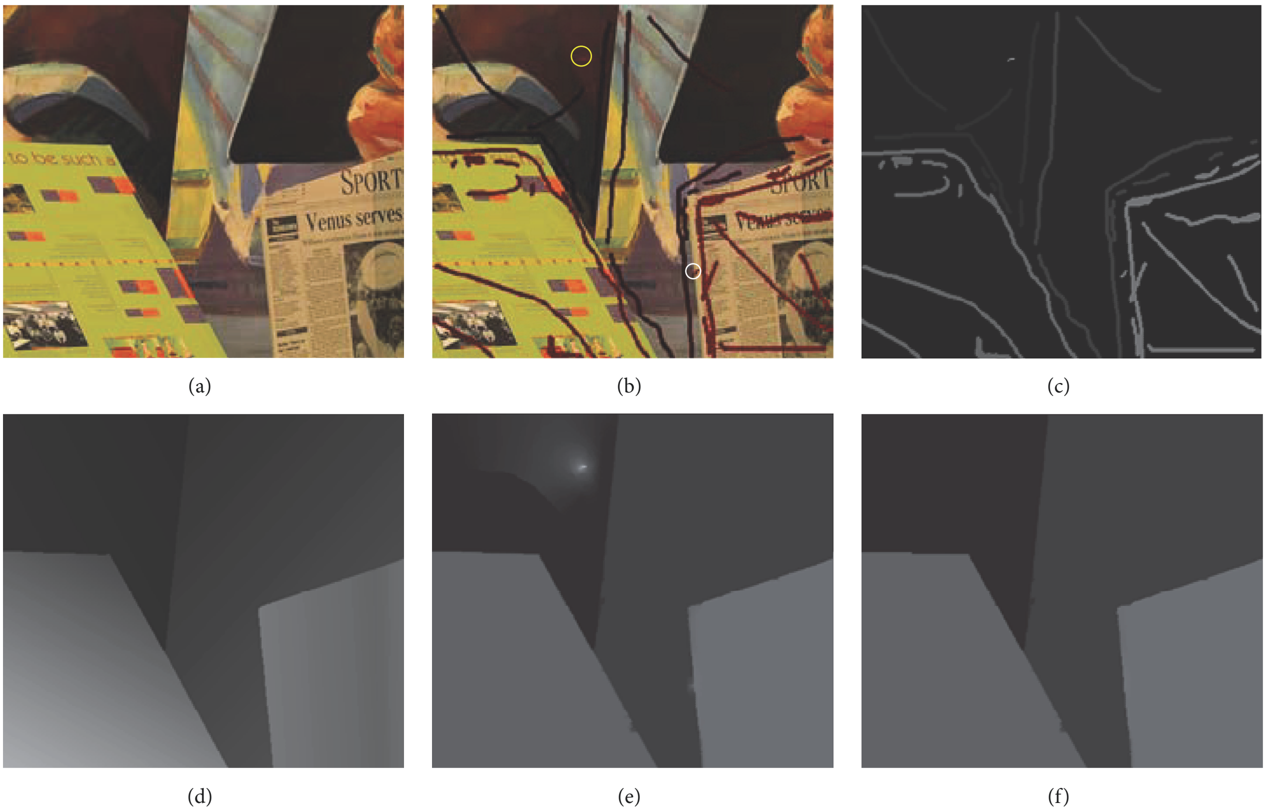


Figure 3: Depths obtained by minimizing quadratic and Welsch M-estimator data fidelity based objective functions, where (a) is input image, (b) is user labeled image (erroneous scribbles at object boundaries are marked by white circles, and errors inside objects or background are marked by yellow circles), (c) is sparse depth map obtained from (b), (d) is ground truth depth, (e) is dense depth map generated by minimizing the quadratic data fidelity based objective function, and (f) is dense depth map generated by minimizing the Welsch M-estimator data fidelity based objective function.

The idea of GIRLS is to determine an upper bound quadratic function and then iteratively minimize the quadratic approximations to obtain a local minimum.

The quadratic upper bound of Welsch function can be obtained by [24]

$$\varphi_{\eta}(x) \le \varphi_{\eta}(y) + \left(1 - \eta\varphi_{\eta}(y)\right)\left(x^{2} - y^{2}\right), \tag{2}$$

with equality only if $x = y$.

Thus the quadratic upper bound of formula (1) is given by

$$\begin{aligned} &\varepsilon_{\eta_1;\eta_2}\left(d; d^{k}\right) \\ &\quad = \sum_{i\in\Omega} e^{-\eta_1 (d_i^k - u_i)^2} \left(d_i - u_i\right)^2 \\ &\qquad + \lambda \sum_{i=1}^{n} \sum_{j\in\mathcal{N}_i} w_{ij} e^{-\eta_2 (d_i^k - d_j^k)^2} \left(d_i - d_j\right)^2 + c, \end{aligned} \tag{3}$$

where c is a constant term which has nothing to do with d and will be ignored in the solving process, d^k denotes estimated depth map at kth iteration, and d_i^k represents its value of the pixel i.

Then, GIRLS for minimizing (1) is to iteratively solve the following problem:

$$d^{k+1} = \arg\min_{d} \varepsilon_{\eta_1;\eta_2}\left(d; d^{k}\right). \tag{4}$$

Let $\mathbf{u} = [u_i]_{n\times 1}$ and $\mathbf{d}^{k+1} = [d_i^{k+1}]_{n\times 1}$; the problem in (4) can be solved in a matrix form as follows:

$$\mathbf{d}^{k+1} = \left(\mathbf{M}^{k} + \lambda \mathbf{L}^{k}\right)^{-1} \mathbf{M}^{k} \mathbf{u}, \tag{5}$$

where $\mathbf{M}^k$ is an $n \times n$ diagonal matrix with ith diagonal entry m_{ii}^k defined in (6), $\mathbf{L}^k = \mathbf{\Lambda}^k - \mathbf{A}^k$ represents the $n \times n$ Laplacian matrix at kth iteration, where $\mathbf{A}^k$ denotes an $n \times n$ affinity matrix with entry a_{ij}^k of the ith row and jth column defined in (7), and $\mathbf{\Lambda}^k$ is an $n \times n$ diagonal matrix with ith diagonal entry Λ_{ii}^k defined in (8).

$$m_{ii}^{k} = \begin{cases} e^{-\eta_1 (d_i^k - u_i)^2} & \text{if } i \in \Omega, \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

$$a_{ij}^{k} = \begin{cases} w_{ij} e^{-\eta_2 (d_i^k - d_j^k)^2} & \text{if } j \in \mathcal{N}_i, \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

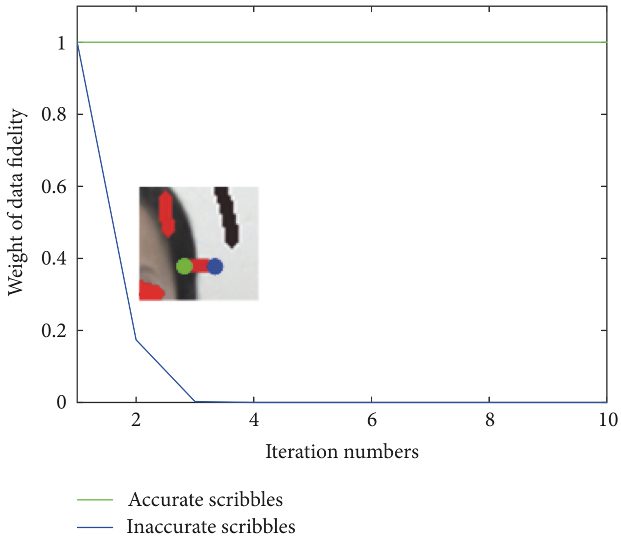


Figure 4: Change curve for weight of data fidelity during iterative solution process where green and blue curves are for labels marked by green and blue circles, respectively.

$$\Lambda_{ii}^{k} = \sum_{j\in\mathcal{N}_i} w_{ij} e^{-\eta_2 (d_i^k - d_j^k)^2}. \qquad (8)$$

In summary, the whole procedure used to minimize (1) is illustrated as follows.

Algorithm 1. The GIRLS algorithm for Welsch data fidelity based sparse-to-dense depth propagation is as follows:

```
(1) Initialization: give parameters $\eta_1, \eta_2, \mu, \lambda, k_{\max}$;
    initialize $k = 0$, $\mathbf{d}^0 = \mathbf{u}$.
    while $k < k_{\max}$ do
(2) Calculate the diagonal entries of $\mathbf{M}^k$ and $\mathbf{\Lambda}^k$ by formula (6) and (8) respectively.
(3) Update the entries of $\mathbf{A}^k$ by formula (7).
(4) Update the Laplacian matrix with $\mathbf{L}^k = \mathbf{\Lambda}^k - \mathbf{A}^k$.
(5) Solve for $\mathbf{d}^{k+1}$ by formula (5).
(6) $k = k + 1$.
    end while.
    Final estimated dense depth $\mathbf{d} = \mathbf{d}^k$.
```

Here, $k_{\max}$ denotes the maximal number of iterations.

2.3. Analysis. Looking at formula (3), we can observe that the data fidelity will be weighted by a Gaussian function of differences between the latest estimated and user input depth values.

At erroneous input regions, inaccurate scribbles will let their depth values be different from neighboring pixels; thus smoothness imposed by regularizer will make estimated depth deviate from user provided depth, and the weight of data fidelity will be decreased to zero during the iteratively solving process. Therefore, the proposed model can suppress inaccurate user scribbles.

At accurate input regions, the depth values of labeled pixels will be consistent with their neighbors; thus the result mainly relies on data fidelity term which makes estimated depth approach user assigned depth. Therefore, the weight of data fidelity will approach 1 during the iterative solution process, and the accurate user scribbles will not be affected by the proposed model.

Figure 4 illustrates the change curve for the data fidelity weight of an input image. We can see that the fidelity weight at erroneous input regions rapidly drops to 0 and it is close to 1 at accurate labeled regions.

3. Experiments

In this section, we report experiments on sparse-to-dense depth propagation for 2D-to-3D conversion with nine representative images, RGBZ_01-09, which are from the RGBZ dataset [34]. Our method was compared to the state of the art: RW [11], hybrid graph-cuts and random walks (HGCRW) [12], nonlocal random walks (NRW) [14], soft segmentation constrained optimization (SCO) [15], OCP [23], and joint static and dynamic guided filtering (SDF) [24]. The experiments were performed on a PC with Intel Core i7 Quad Processor (4 GHz) (more experimental results and source code are available at https://github.com/tcyhx/rme_2dto3d).

As the quantitative evaluation metric, we used structural similarity (SSIM) [35] since it can predict human perception of image quality. Similar to Konno et al. [36], the standard deviation of Gaussian function in SSIM was set to 4 so that it can evaluate the similarity of semiglobal structure. The higher SSIM value shows a better performance.

3.1. Choice of the Parameters. The parameters λ, μ, η_1, η_2, and $k_{\max}$ should be set to begin with our sparse-to-dense depth propagation algorithm. The parameter λ is used to balance data fidelity with regularizer and has impact on

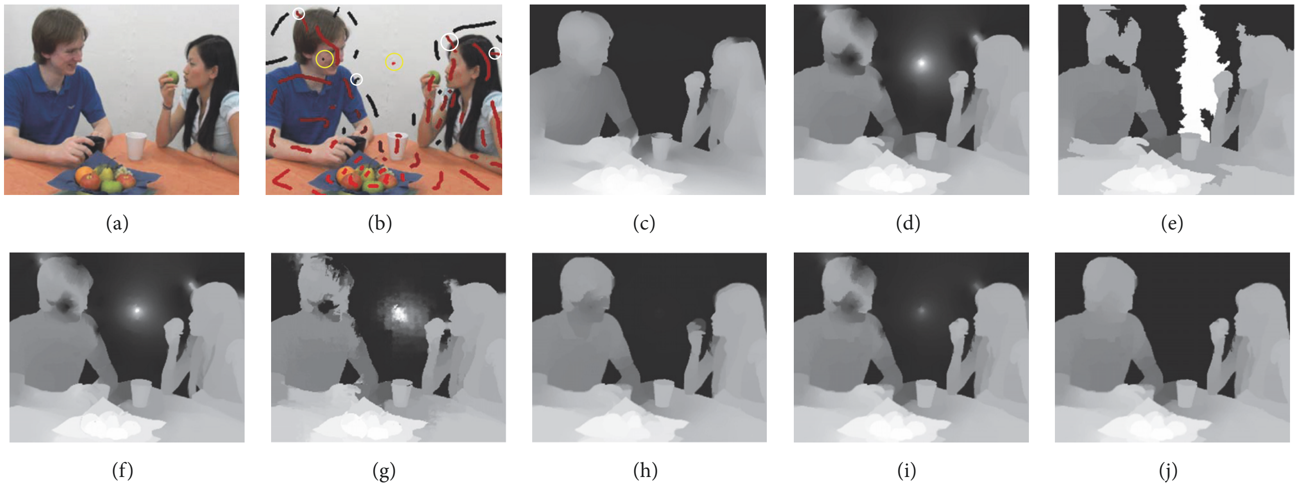


Figure 5: Results of different methods on RGBZ_01 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

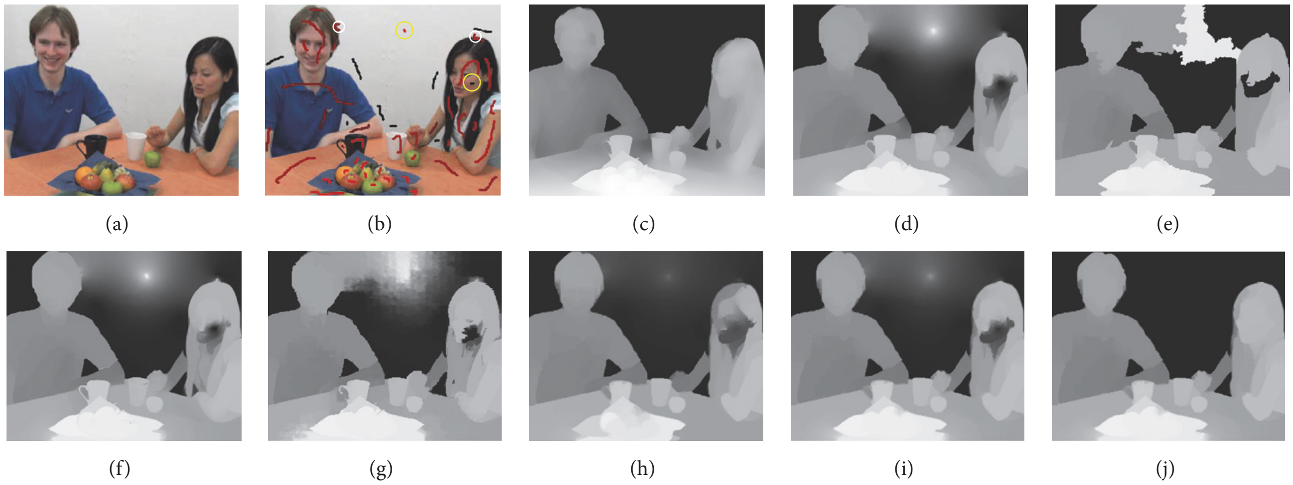


Figure 6: Results of different methods on RGBZ_02 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

depth smoothness. The bandwidth parameters μ and η_2 are utilized to adjust the performance of depth discontinuities preservation. Liu et al. [37] have proposed an adaptive method to calculate the bandwidth according to the local depth smoothness. The parameter η_1 has influence on the strength of resistance to outliers. $k_{\max}$ is used to terminate iterations. Our algorithm typically converges in less than 10 iterations. Thus $k_{\max}$ is fixed to 10 in our method. We find that the choice for $\lambda = 10$, $\mu = 2000$, $\eta_1 = 1000$, and $\eta_2 = 0.1$ is proper for most cases.

3.2. Comparison with Existing Methods in the Presence of Erroneous Scribbles. In this subsection, we roughly draw labels across some randomly selected object boundaries, and these erroneous labeled regions are marked by white circles in Figures 5(b), 6(b), 7(b), 8(b), 9(b), 10(b), 11(b), 12(b), and 13(b). We also randomly add erroneous scribbles inside objects or background which are marked by yellow circles in Figures 5(b), 6(b), 7(b), 8(b), 9(b), 10(b), 11(b), 12(b), and 13(b).

Table 1 shows quantitative comparisons in the presence of erroneous scribbles. It can be seen from Table 1 that the proposed method achieves the best performance for all scenes in terms of the SSIM.

Figures 5–13 show visual comparisons for estimated depth when erroneous scribbles are present. The RW [11] algorithm does not change user provided labels during the solution process. Therefore, it cannot obtain correct depth values around the inaccurate labels (see Figures 5(d), 6(d), 7(d), 8(d), 9(d), 10(d), 11(d), 12(d), and 13(d)). The HGCRW [12] attempts to enhance the depth discontinuities of RW using hard constraints provided by GC segmentation, but

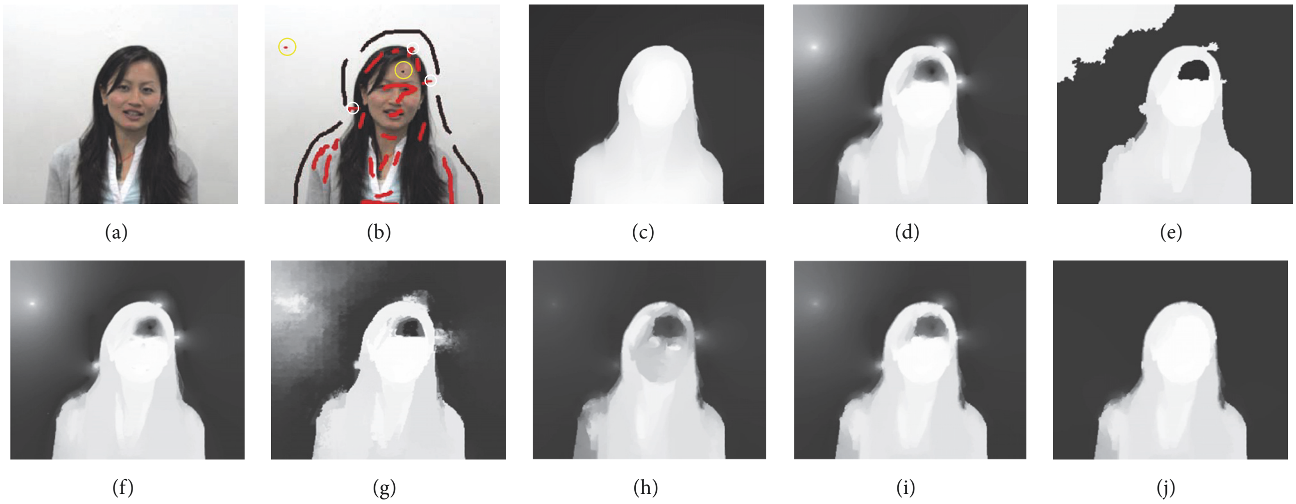


Figure 7: Results of different methods on RGBZ_03 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

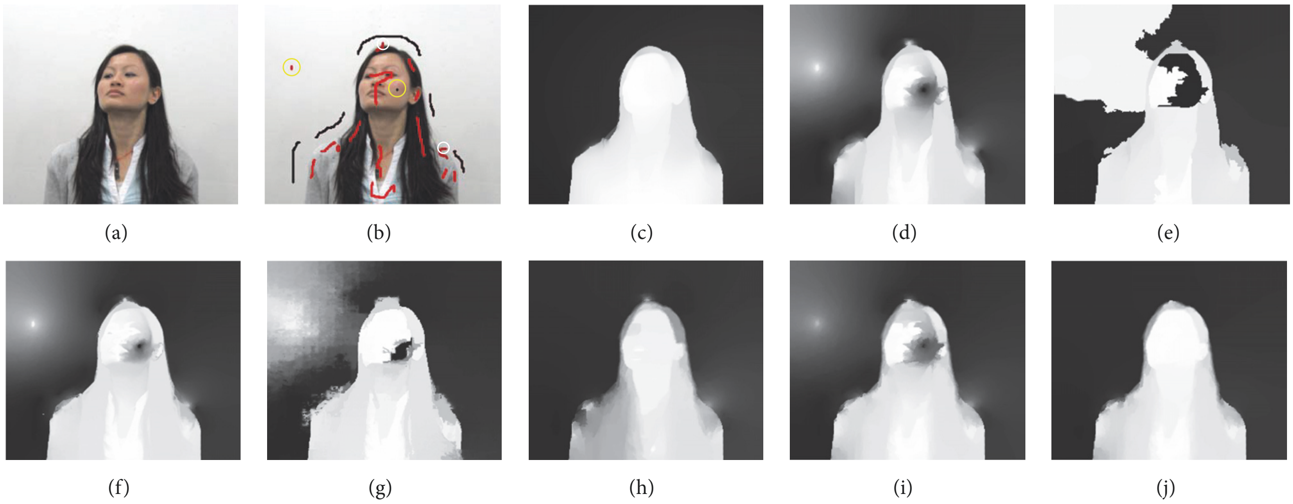


Figure 8: Results of different methods on RGBZ_04 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

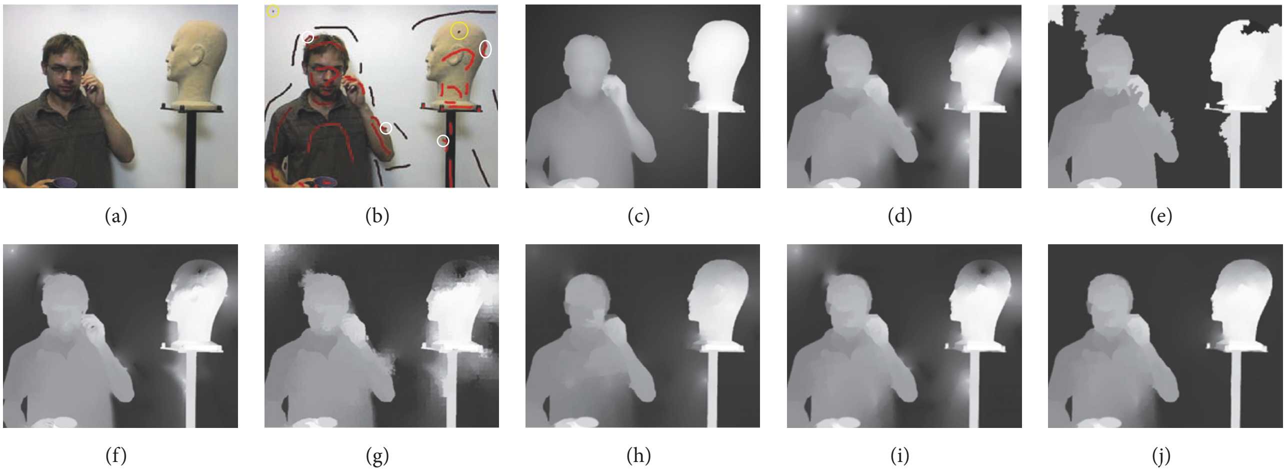


Figure 9: Results of different methods on RGBZ_05 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

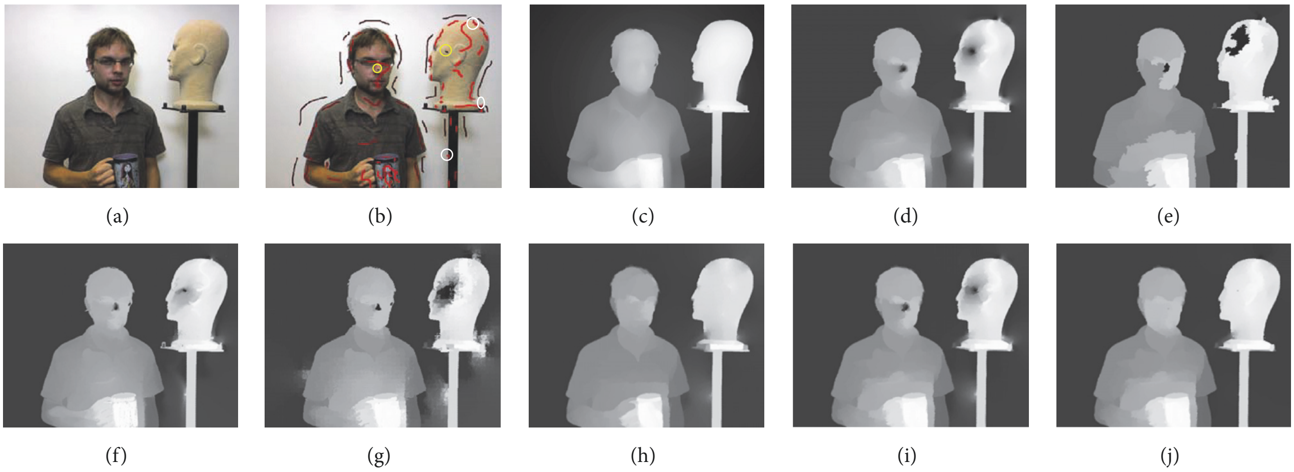


FIGURE 10: Results of different methods on RGBZ_06 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

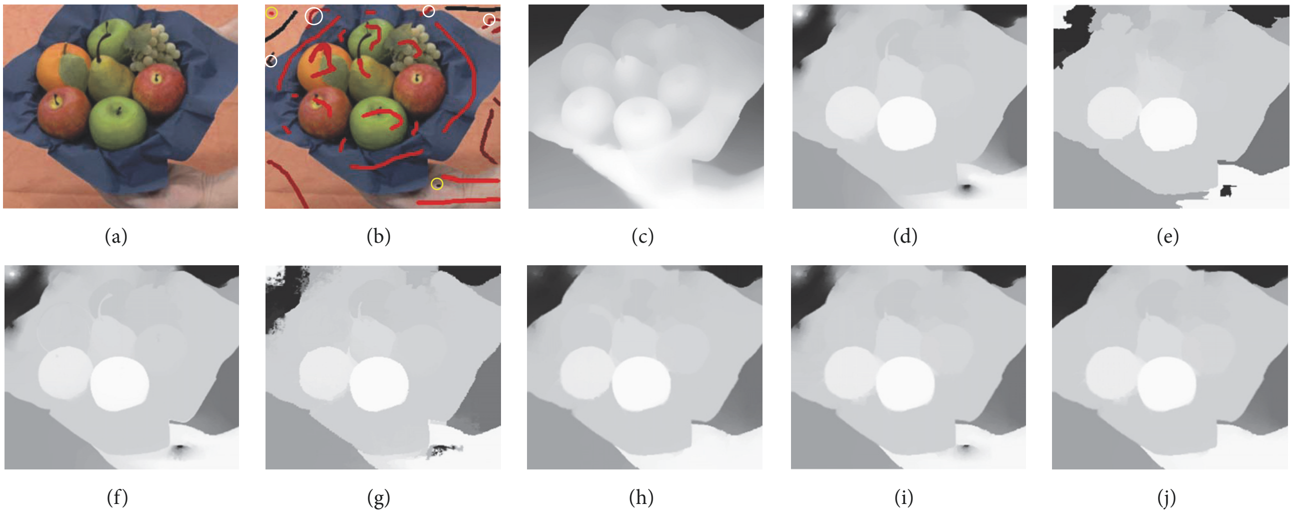


FIGURE 11: Results of different methods on RGBZ_07 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

TABLE 1: SSIM comparison in the presence of erroneous input. The first and second best SSIM at each row are shown in bold and italic, respectively.

Images	Methods						
	RW [11]	HGCRW [12]	NRW [14]	SCO [15]	OCP [23]	SDF [24]	Ours
RGBZ_01	0.69	0.69	0.71	0.72	*0.84*	0.72	**0.89**
RGBZ_02	0.66	*0.73*	0.67	0.66	0.65	0.66	**0.88**
RGBZ_03	0.70	*0.77*	0.73	0.67	0.74	0.75	**0.86**
RGBZ_04	0.70	0.68	0.73	0.66	*0.85*	0.75	**0.91**
RGBZ_05	0.78	0.78	0.81	0.74	*0.86*	0.81	**0.89**
RGBZ_06	0.84	0.82	0.84	0.78	*0.85*	0.83	**0.88**
RGBZ_07	0.83	0.81	0.82	0.80	*0.85*	0.83	**0.86**
RGBZ_08	0.86	0.85	0.86	0.83	**0.91**	*0.87*	**0.91**
RGBZ_09	0.79	0.74	*0.81*	0.67	*0.81*	*0.81*	**0.92**
Average	0.76	0.76	0.78	0.73	*0.82*	0.78	**0.89**

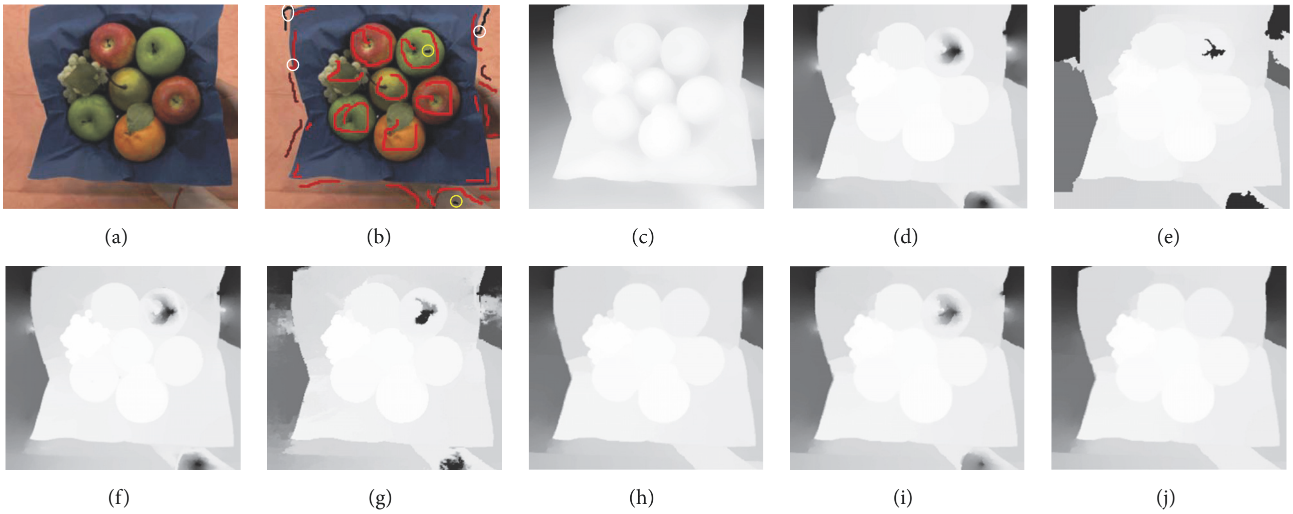


Figure 12: Results of different methods on RGBZ_08 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

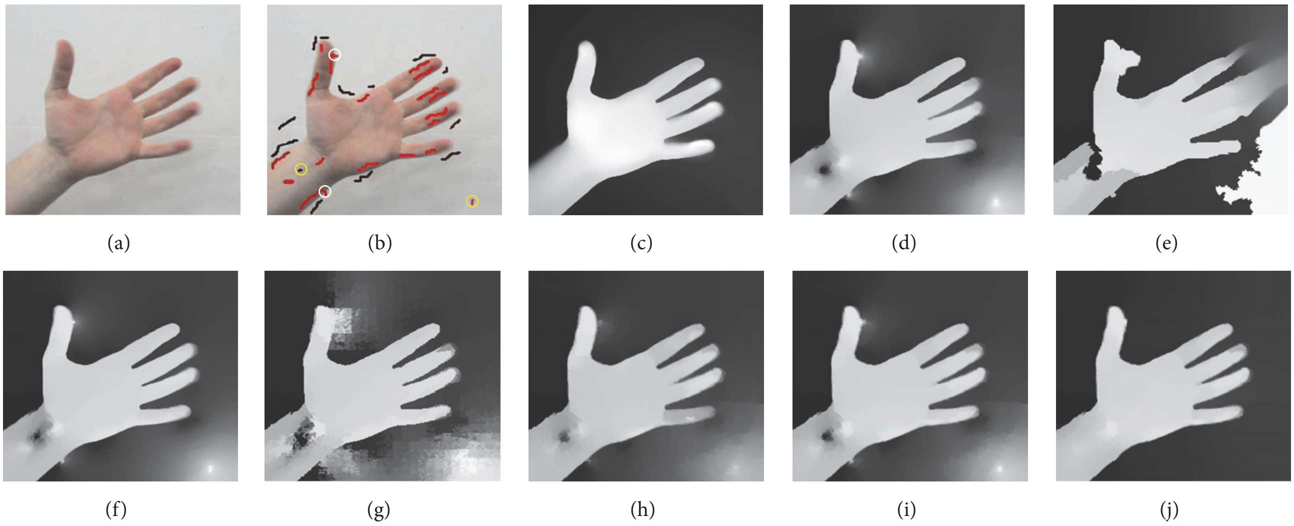


Figure 13: Results of different methods on RGBZ_09 in the presence of erroneous scribbles. (a) is input image. (b) is user labeled image (scribbles inside white and yellow circles are inaccurate). (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

incorrect segmentation caused by erroneous scribbles introduces serious depth artifacts (see Figures 5(e), 6(e), 7(e), 8(e), 9(e), 10(e), 11(e), 12(e), and 13(e)). The NRW [14] utilizes the nonlocal pairwise constraints to improve depth quality, but this cannot suppress inaccurate labels (see Figures 5(f), 6(f), 7(f), 8(f), 9(f), 10(f), 11(f), 12(f), and 13(f)). The SCO [15] employs depth consistency between pixels and superpixels to preserve depth boundaries, but may propagate erroneous scribbles to more regions (see Figures 5(g), 6(g), 7(g), 8(g), 9(g), 10(g), 11(g), 12(g), and 13(g)). The OCP [23] alleviates the influence of inaccurate scribbles using global and local color distribution underlying the user provided scribbles, but still cannot suppress some erroneous scribbles (see Figures 5(h), 6(h), 7(h), 8(h), 9(h), 10(h), 11(h), 12(h), and 13(h)). The SDF [24] uses the Welsch function for the regularizer to suppress depth artifacts caused by structural differences between color and depth images, but cannot handle erroneous labels (see Figures 5(i), 6(i), 7(i), 8(i), 9(i), 10(i), 11(i), 12(i), and 13(i)). Thanks to the Welsch M-estimator data fidelity, our method suppresses erroneous input successfully and obtains robust depth estimation results in the presence of erroneous scribbles (see Figures 5(j), 6(j), 7(j), 8(j), 9(j), 10(j), 11(j), 12(j), and 13(j)).

3.3. Comparison with Existing Methods in the Absence of Erroneous Scribbles. In this subsection, we perform experiments on depth estimation with human interactions for 2D-to-3D

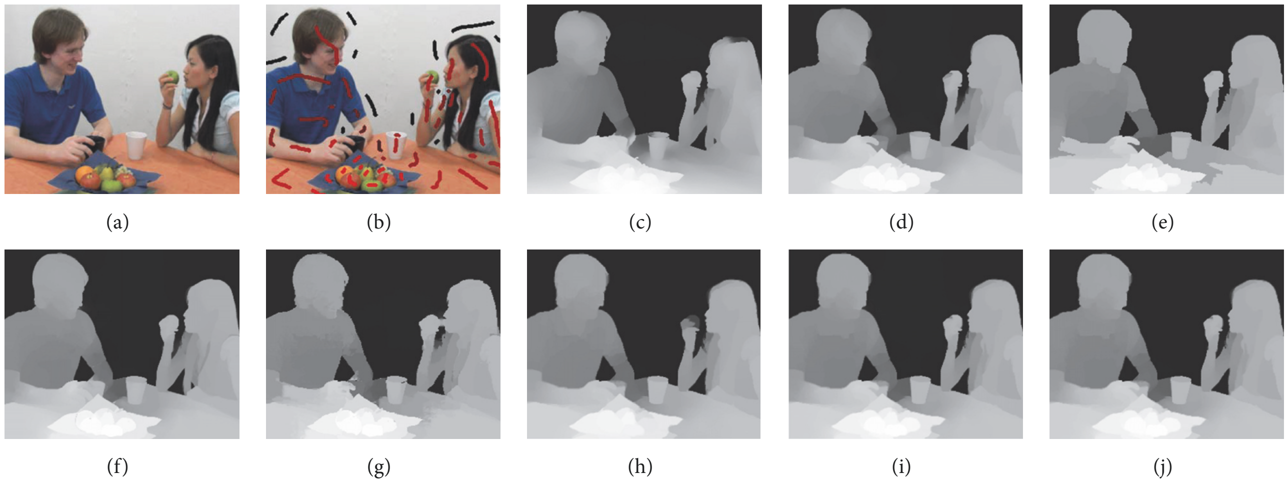


Figure 14: Results of different methods on RGBZ_01 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

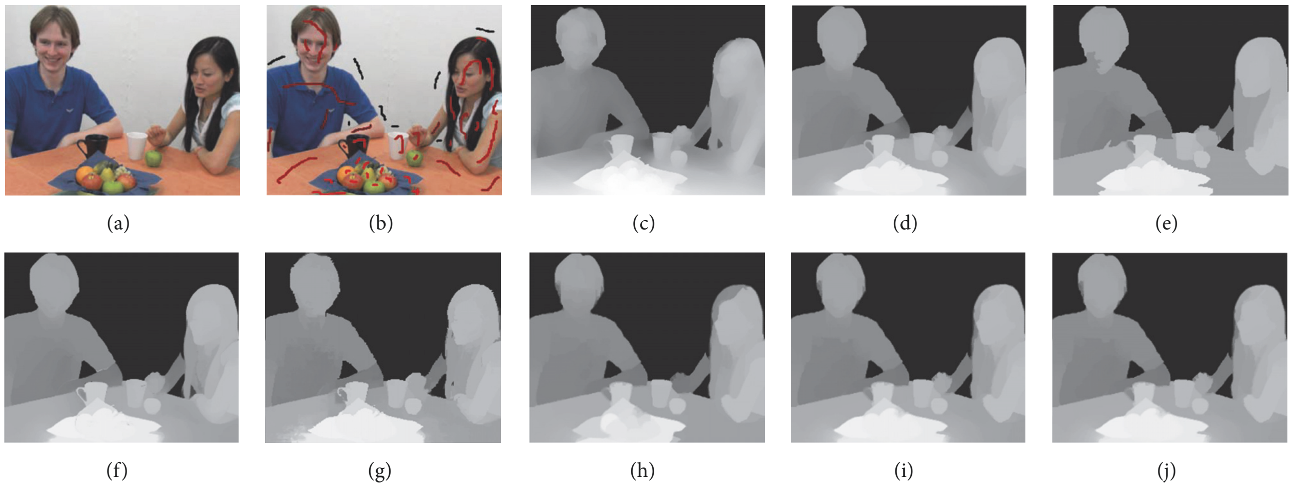


Figure 15: Results of different methods on RGBZ_02 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

conversion when erroneous input is absent. Table 2 presents SSIM comparisons for estimated depth. Visual comparisons are shown in Figures 14–22.

In Table 2, it shows that our method has the same performance as SDF [24] in terms of SSIM. The reason is that the proposed method generates weights approaching 1 for data fidelity in the absence of erroneous input, which has been illustrated in Figure 4. From the data in Table 2, we can find that our method has the second highest SSIM in average. Therefore, the proposed method has comparable performance to the state of the art approaches in the absence of erroneous scribbles.

From the above experiments, we can see that the proposed method outperforms the state of the art methods both qualitatively and quantitatively in the presence of inaccurate scribbles. In addition, our method shows comparable performance when accurate scribbles are provided. Therefore, the proposed method can be used in depth estimation problems for 2D-to-3D conversion under various cases.

4. Conclusion and Future Work

We propose a robust sparse-to-dense depth propagation method for images or key frames in semiautomatic 2D-to-3D conversion. Depth estimation is formulated as a nonconvex problem. We leverage the Welsch M-estimator to construct data fidelity term, and exploit the outlier resistance property of redescending M-estimator to suppress erroneous scribbles. The experiments demonstrate that our method is more robust than the state of the art methods when inaccurate input is present, and obtains comparable performance in the absence of erroneous scribbles.

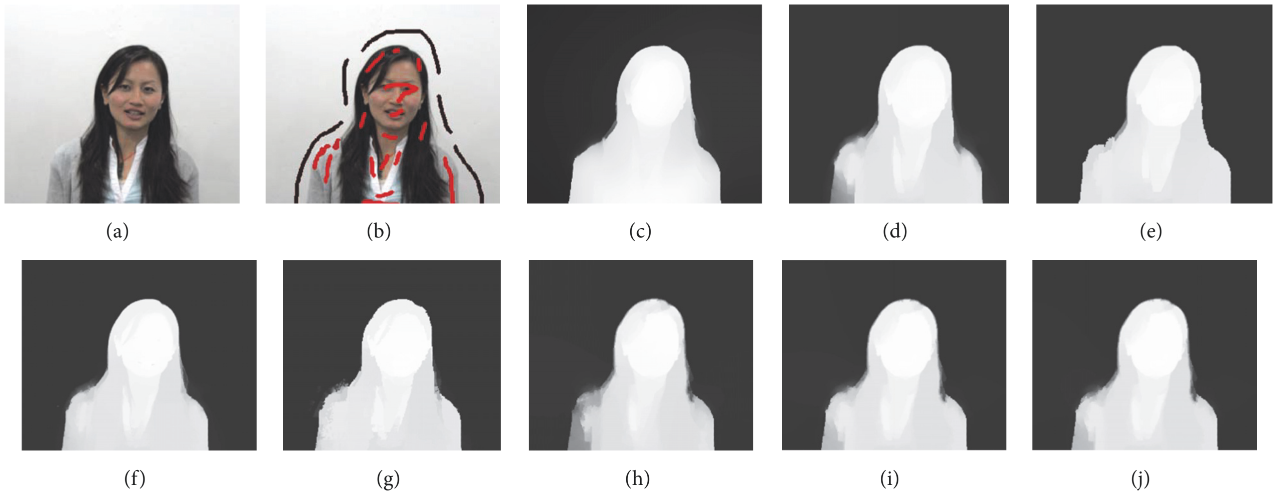


Figure 16: Results of different methods on RGBZ_03 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

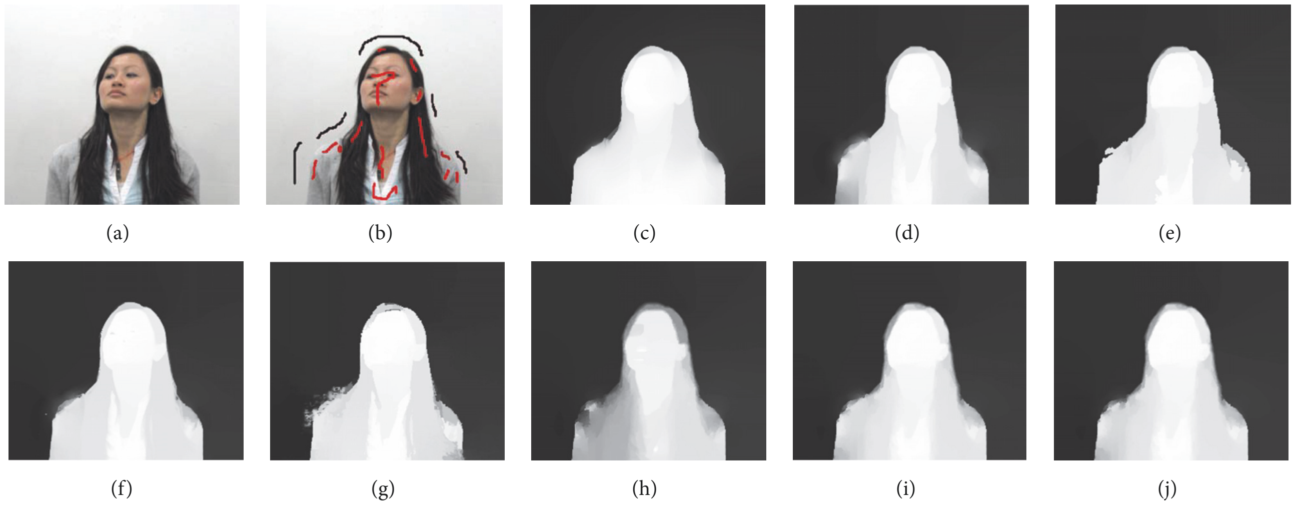
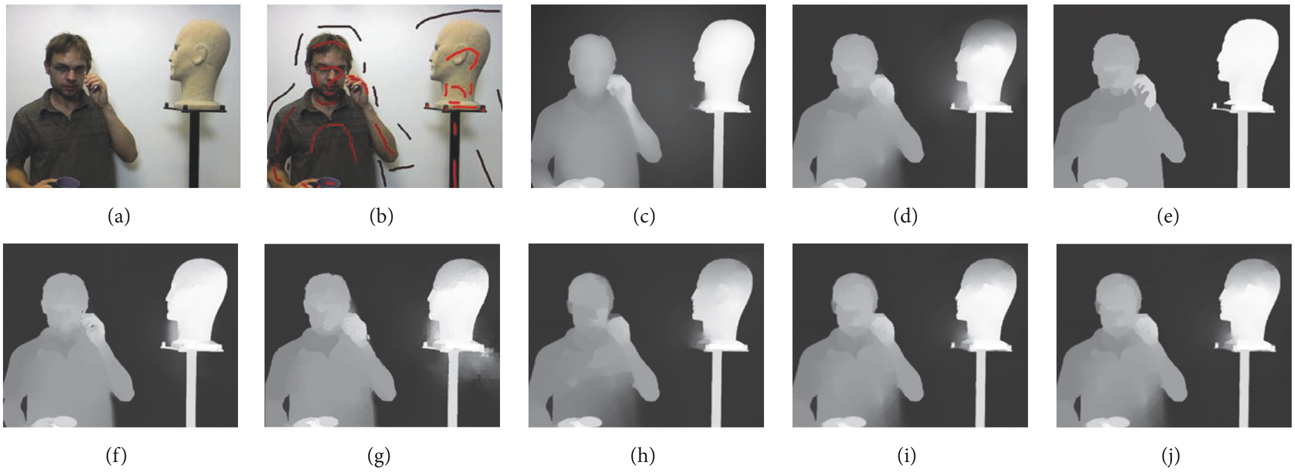


Figure 17: Results of different methods on RGBZ_04 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

Table 2: SSIM comparison in the absence of erroneous input. The first and second best SSIM at each row are shown in bold and italic, respectively.

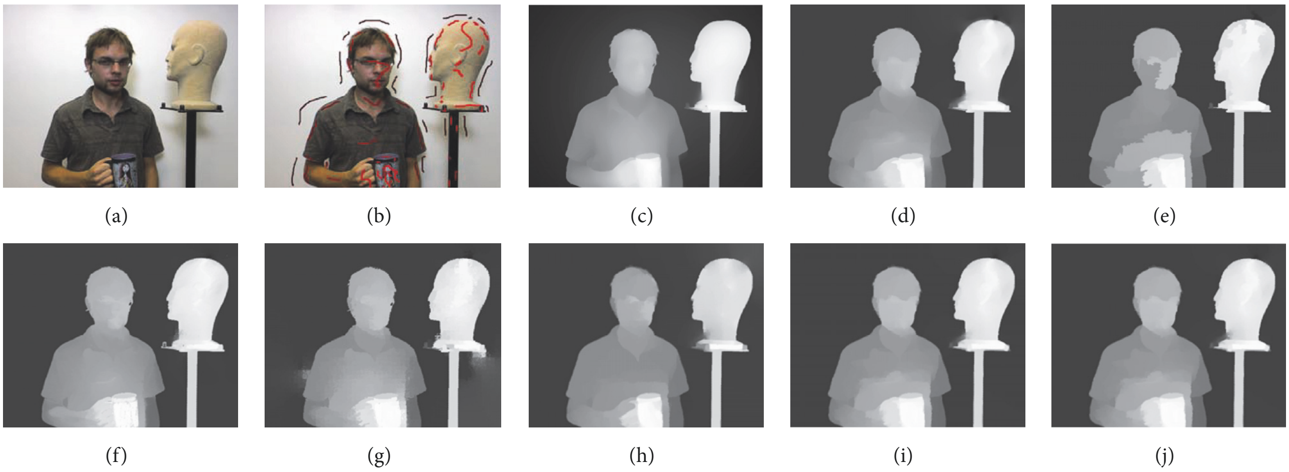
Images	Methods						
	RW [11]	HGCRW [12]	NRW [14]	SCO [15]	OCP [23]	SDF [24]	Ours
RGBZ_01	*0.89*	0.87	**0.90**	0.87	*0.89*	**0.90**	**0.90**
RGBZ_02	**0.89**	0.85	**0.89**	0.86	0.86	*0.87*	*0.87*
RGBZ_03	*0.86*	**0.87**	**0.87**	0.85	0.85	*0.86*	*0.86*
RGBZ_04	0.90	*0.91*	**0.92**	0.87	0.88	*0.91*	*0.91*
RGBZ_05	0.88	*0.89*	**0.90**	0.86	0.88	*0.89*	*0.89*
RGBZ_06	**0.89**	0.85	**0.89**	0.85	0.86	*0.88*	*0.88*
RGBZ_07	**0.86**	*0.84*	**0.86**	*0.84*	**0.86**	**0.86**	**0.86**
RGBZ_08	**0.92**	0.88	**0.92**	0.89	*0.91*	*0.91*	*0.91*
RGBZ_09	*0.92*	0.84	**0.93**	0.91	0.91	*0.92*	*0.92*
Average	*0.89*	0.87	**0.90**	0.87	0.88	*0.89*	*0.89*



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

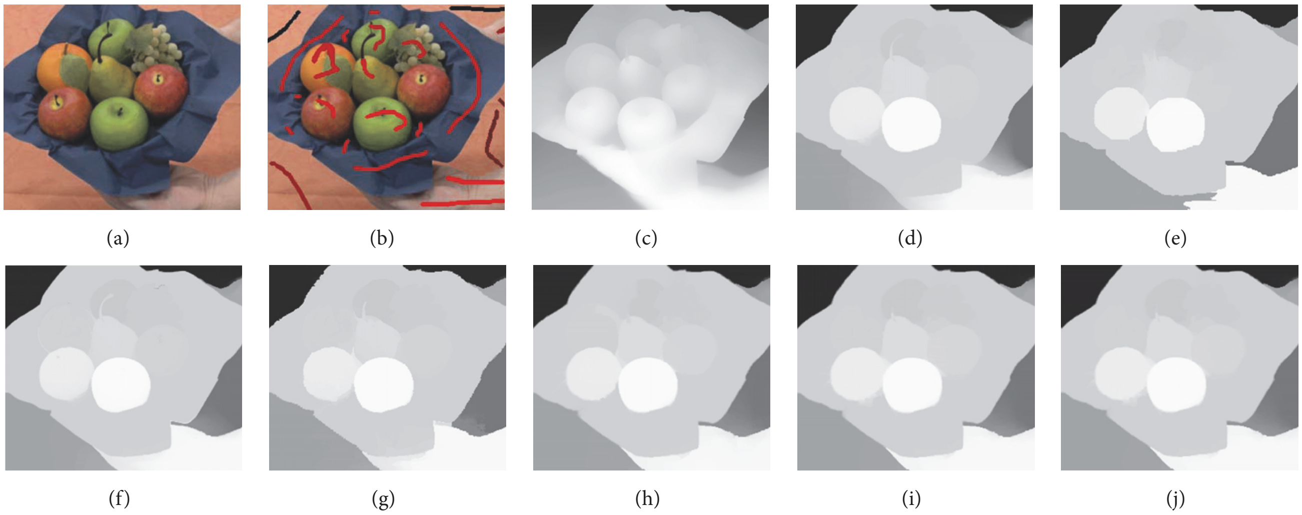
Figure 18: Results of different methods on RGBZ_05 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 19: Results of different methods on RGBZ_06 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 20: Results of different methods on RGBZ_07 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

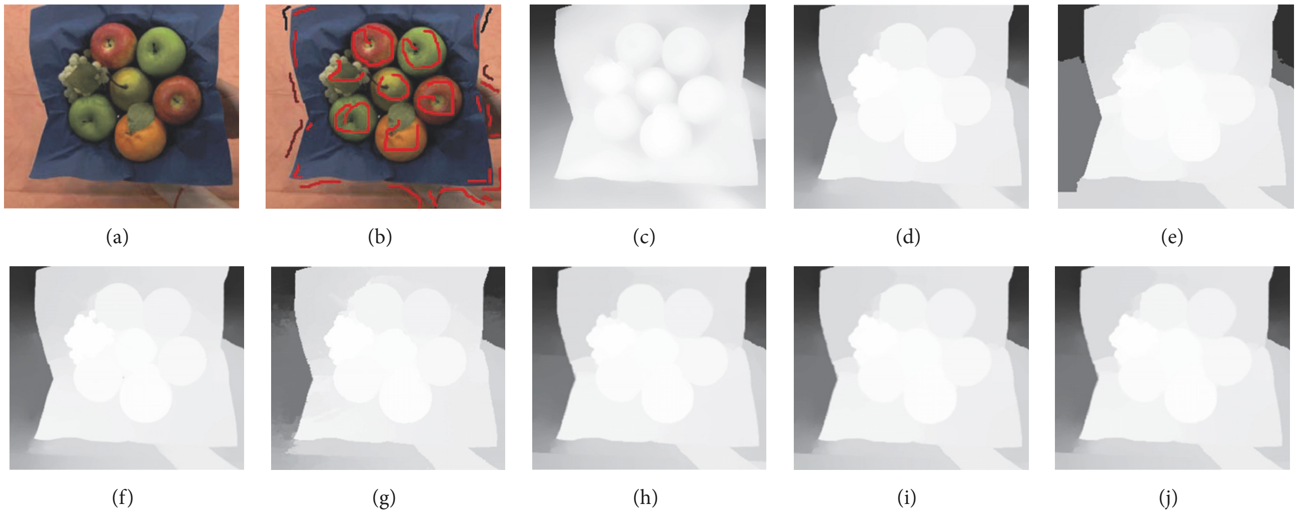

(a) (b) (c) (d) (e)
(f) (g) (h) (i) (j)

Figure 21: Results of different methods on RGBZ_08 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

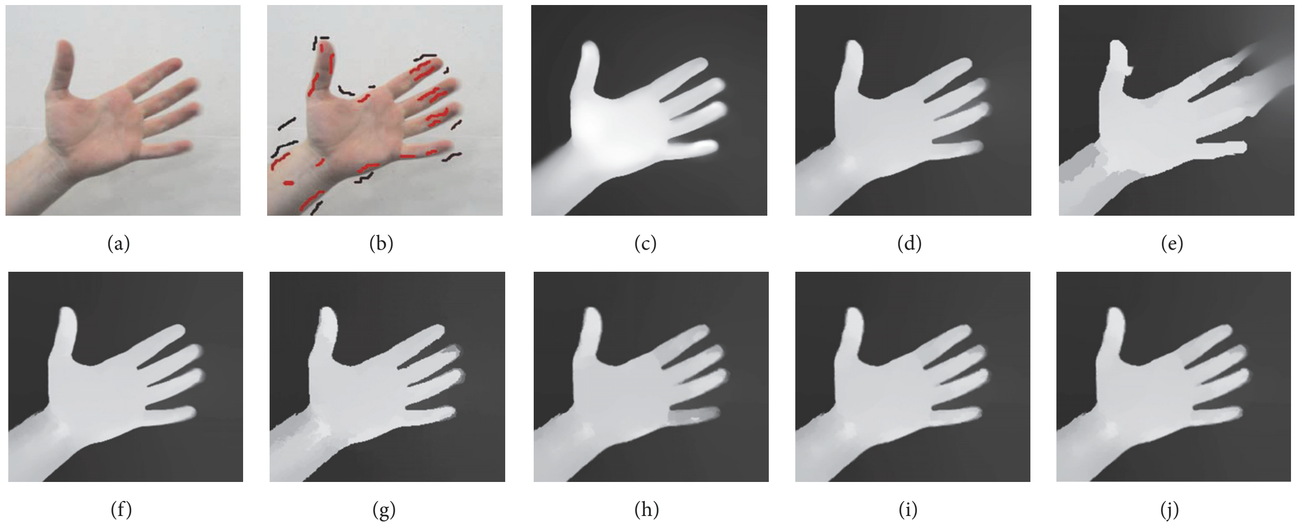

(a) (b) (c) (d) (e)
(f) (g) (h) (i) (j)

Figure 22: Results of different methods on RGBZ_09 in the absence of erroneous scribbles. (a) is input image. (b) is user labeled image. (c) is ground truth depth. (d) is result of RW. (e) is result of HGCRW. (f) is result of NRW. (g) is result of SCO. (h) is result of OCP. (i) is result of SDF. (j) is result of the proposed method.

The parameters of our method are set empirically. In the future, an optimal parameters setting scheme according to depth properties should be proposed. In addition, we will apply our method to perform depth propagation from key frames to nonkey frames.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant nos. LY16F010014 and LY18F020025, the Educational Commission of Zhejiang Province of China under Grant no. Y201533511, the National Natural Science Foundation of China under Grant nos. 61502256 and 61671260, and Ningbo Natural Science Foundation under Grant no. 2017A610109.

References

[1] Y. Feng, J. Ren, and J. Jiang, "Object-based 2D-to-3D video conversion for effective stereoscopic content generation in 3D-TV applications," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 500–509, 2011.

[2] R. Rzeszutek, R. Phan, and D. Androutsos, "Depth estimation for semi-automatic 2D to 3D conversion," in *Proceedings of the 20th ACM International Conference on Multimedia*, pp. 817–820, ACM, November 2012.

[3] 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.

[4] 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 the Eighth Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pp. 47–54, ACM, August 2011.

[5] J. Xie, R. Girshick, and A. Farhadi, “Deep3D: fully automatic 2D-to-3D video conversion with deep convolutional neural networks,” *European Conference on Computer Vision*, pp. 842–857, 2016.

[6] R. Garg, B. G. Vijay Kumar, G. Carneiro, and I. Reid, “Unsupervised CNN for single view depth estimation: geometry to the rescue,” *European Conference on Computer Vision*, pp. 740–756, 2016.

[7] A. Grigorev, F. Jiang, S. Rho, W. J. Sori, S. Liu, and S. Sai, “Depth estimation from single monocular images using deep hybrid network,” *Multimedia Tools and Applications*, pp. 1–20, 2016.

[8] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, “Deeper depth prediction with fully convolutional residual networks,” in *Proceedings of the 4th International Conference on 3D Vision, 3DV 2016*, pp. 239–248, IEEE, October 2016.

[9] C. Godard, O. M. Aodha, and G. J. Brostow, “Unsupervised monocular depth estimation with left-right consistency,” 2016, https://arxiv.org/abs/1609.03677.

[10] Y. Kuznietsov, J. Stückler, and B. Leibe, “Semi-supervised deep learning for monocular depth map prediction,” 2017, https://arxiv.org/abs/1702.02706.

[11] R. Rzeszutek, R. Phan, and D. Androutsos, “Semi-automatic synthetic depth map generation for video using random walks,” in *Proceedings of the 2011 12th IEEE International Conference on Multimedia and Expo, ICME 2011*, 6, 1 pages, IEEE, July 2011.

[12] R. Phan and D. Androutsos, “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.

[13] C. Rother, V. Kolmogorov, and A. Blake, “GrabCut: interactive foreground extraction using iterated graph cuts,” *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, 2004.

[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] 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.

[16] A. Lopez, E. Garces, and D. Gutierrez, “Depth from a single image through user interaction,” in *CEIG*, pp. 11–20, 2014.

[17] M. Becker, M. Baron, D. Kondermann, M. Bußler, and V. Helzle, “Movie dimensionalization via sparse user annotations,” in *3DTV-Conference: The True Vision-Capture, Transmission and Dispaly of 3D Video (3DTV-CON)*, 2013.

[18] A. Kawai and N. Sasaki, “A depth creation technique by modifying unicursal outlines for 2D/3D conversion,” in *Proceedings of 3DSA*, vol. 5, p. 3, 2013.

[19] D. Donatsch, N. Färber, and M. Zwicker, “3D conversion using vanishing points and image warping,” in *Proceedings of the 2013 3DTV-Conference: The True Vision-Capture, Transmission and Dispaly of 3D Video, 3DTV-CON 2013*, pp. 1–4, IEEE, October 2013.

[20] Z. Zhang, C. Zhou, Y. Wang, and W. Gao, “Interactive stereoscopic video conversion,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 10, pp. 1795–1808, 2013.

[21] S. Iizuka, Y. Endo, Y. Kanamori, J. Mitani, and Y. Fukui, “Efficient Depth Propagation for Constructing a Layered Depth Image from a Single Image,” in *Computer Graphics Forum*, vol. 33, pp. 279–288, Wiley Online Library, 2014.

[22] J. Liao, S. Shen, and E. Eisemann, “Depth annotations: Designing depth of a single image for depth-based effects,” *Proceedings of Graphics Interface (GI)*, 2017.

[23] 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.

[24] B. Ham, M. Cho, and J. Ponce, “Robust image filtering using joint static and dynamic guidance,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015*, pp. 4823–4831, June 2015.

[25] K. Subr, S. Paris, C. Soler, and J. Kautz, “Accurate binary image selection from inaccurate user input,” in *Computer Graphics Forum*, vol. 32, pp. 41–50, Wiley Online Library, 2013.

[26] J. Bai and X. Wu, “Error-tolerant scribbles based interactive image segmentation,” in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 392–399, June 2014.

[27] Y. Yang, Y. Feng, and J. A. Suykens, “Robust low-rank tensor recovery with regularized redescending M-estimator,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 9, pp. 1933–1946, 2016.

[28] Y. Kim, S. Choi, C. Oh, and K. Sohn, “A majorize-minimize approach for high-quality depth upsampling,” in *Proceedings of the IEEE International Conference on Image Processing, ICIP 2015*, pp. 392–396, September 2015.

[29] Y. Kim, B. Ham, C. Oh, and K. Sohn, “Structure selective depth superresolution for RGB-D cameras,” *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5227–5238, 2016.

[30] W. Liu, S. Jia, P. Li, X. Chen, J. Yang, and Q. Wu, “An MRF-based depth upsampling: Upsample the depth map with its own property,” *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1708–1712, 2015.

[31] W. Liu, X. Chen, J. Yang, and Q. Wu, “Robust color guided depth map restoration,” *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 315–327, 2017.

[32] N. Bissantz, L. Dümbgen, A. Munk, and B. Stratmann, “Convergence analysis of generalized iteratively reweighted least squares algorithms on convex function spaces,” *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1828–1845, 2008.

[33] Y. Li, G. Li, and Y. Hong-Xing, “A novel method of depth-image-based view synthesis,” *Journal of the Graduate School of the Chinese Academy of Sciences*, vol. 5, pp. 638–644, 2010.

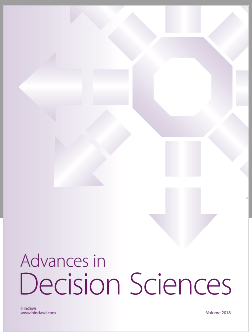
[34] C. Richardt, C. Stoll, N. A. Dodgson, H.-P. Seidel, and C. Theobalt, “Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos,” in *Computer Graphics Forum*, vol. 31, pp. 247–256, Wiley Online Library, 2012.

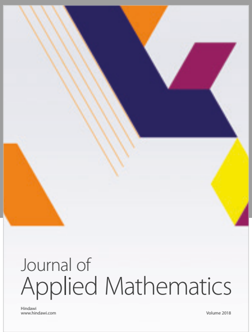
[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[36] Y. Konno, M. Tanaka, M. Okutomi, Y. Yanagawa, K. Kinoshita, and M. Kawade, “Depth map upsampling by self-guided residual interpolation,” in *Proceedings of the 23rd International Conference on Pattern Recognition, ICPR 2016*, pp. 1394–1399, Mexico, December 2016.

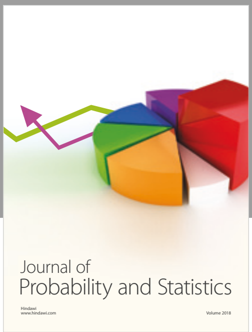
[37] W. Liu, X. Chen, J. Yang, and Q. Wu, "Variable bandwidth weighting for texture copy artifact suppression in guided depth upsampling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2072–2085, 2017.

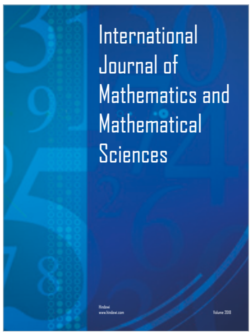
Advances in
Operations Research


Advances in
Decision Sciences

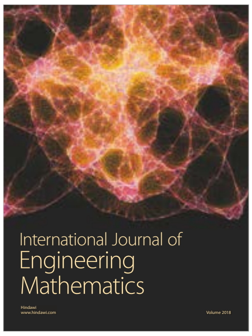

Journal of
Applied Mathematics

The Scientific
World Journal

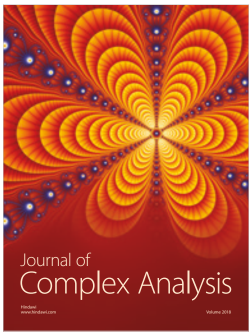

Journal of
Probability and Statistics

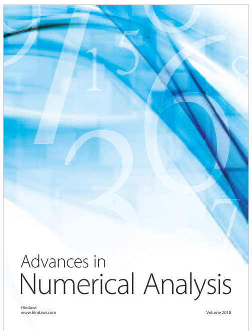

International
Journal of
Mathematics and
Mathematical
Sciences

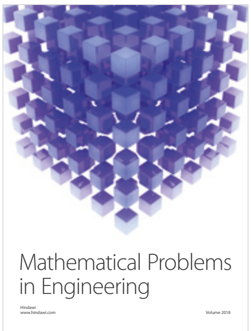
Journal of
Optimization

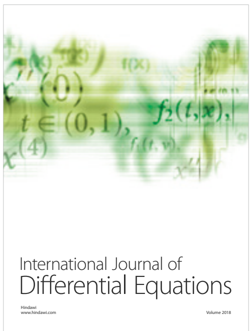

International Journal of
Engineering
Mathematics

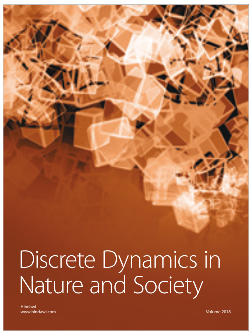
International Journal of
Analysis

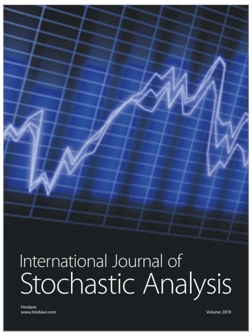

Journal of
Complex Analysis


Advances in
Numerical Analysis


Mathematical Problems
in Engineering

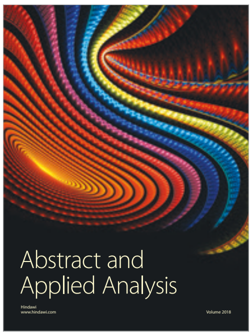

International Journal of
Differential Equations

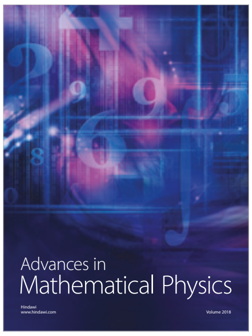

Discrete Dynamics in
Nature and Society


International Journal of
Stochastic Analysis

Journal of
Mathematics

Journal of
Function Spaces


Abstract and
Applied Analysis


Advances in
Mathematical Physics