International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 10, No. 5 (2017), pp.129-142
http://dx.doi.org/10.14257/ijsip.2017.10.5.11

Depth Map Enhancement based on Improved Adaptive Total Generalized Variation for ToF Camera

Lujiang Liu, Gaopeng Zhao* and Yuming Bo

School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China;
zhaogaopeng@njust.edu.cn

Abstract

Time-of-Flight (ToF) cameras have become a promising sensor in several applications as it can directly provide the depth measurements. However, there are two main drawbacks, one is the limited spatial resolution of currently available low-cost ToF sensors; the other is the noise and the edge distortion within the depth measurement. In order to address the shortcomings, a new depth map enhancement method is proposed by fusing a ToF sensor and a coupled 2D high resolution color sensor. In this paper, we incorporate the low resolution depth data with the edge and the gradient information of the high resolution intensity image into the improved second order total generalized variation (TGV) model. Also a parameter adaptive mechanism is designed for the regularization term. Through the optimization solution of the improved TGV model, the high resolution depth map is obtained while the edges and details are enhanced. To investigate the effectiveness of the proposed method, experiments on the public Middlebury datasets and the real sensor datasets are performed. Results show that the proposed method is effective and outperforms the compared methods.

Keywords: *depth map enhancement; total generalized variation; time-of-flight camera; super resolution*

1. Introduction

Depth sensors are widely used in numerous scientific domains in recent years such as autonomous vehicles, human computer interaction, robotics et al [1-4]. The typical depth sensors include passive sensing (stereo vision) and active sensing (structured light, lidar, time-of-flight camera). The accuracy of the depth map affects the subsequent processing and affects the effects of applications especially some high-accuracy applications. Hence the demand for a high quality depth map is one of the active research areas in computer vision.

Time-of-Flight (ToF) camera is a relatively new type of sensor, and it is a very promising sensor as its advantages include registered depth and intensity data at a high frame rate, compact design, low weight and reduced power consumption. However compared to other depth sensors, ToF cameras cannot be considered yet as a mature sensor. The disadvantages are obvious such as low resolution, motion artifacts, edge distortion, ambient light noise, *et al.*, [1,5]. Despite its disadvantages, it is already showing great potential in many applications where fast 3D depth data is needed, such as pose estimation [6,7], obstacle avoidance [8,9] and others.

To obtain the high quality depth map, some scholars have attempted to handle the drawbacks of the ToF cameras. However, up to now, the problem is still an open research area. To increase both the resolution and the quality, a new depth map enhancement method based on improved Total Generalized Variation (TGV) is proposed in the paper by using the sensor fusion. Assuming that the TOF camera and the coupled 2D monocular

ISSN: 2005-4254 IJSIP

camera are calibrated [10,11,12], we only focus on the data processing in this paper, and the calibration of the two sensors is beyond the scope of this article.

In this paper, the accurate textural edge information of intensity data is incorporated into the high order regularization term to handle the depth distortion, which often exists in the object boundary of the ToF depth data. Besides, the parameters in the regularization term are set to the constant value in the most existing works, which is not appropriate and maybe lead to edge blurring and detail loss. So a parameters adaptive mechanism is designed in the paper.

The main contribution of our approach is that by improving the TGV model with edge factor and adaptive parameters, not only the high resolution dense depth map can be obtained, but also the texture edge details can be well enhanced. Also, the parameters adaptive mechanism enhances the applicability of the approach. The experiments with publicly standard datasets and real images demonstrate the effectiveness of the proposed approach. Compared to state of the art methods, the proposed approach shows superior performance on the test datasets.

The rest of the paper is organized in the following sequence. In Section 2 the related works in recent years are described in detail. The TGV based depth upsampling method is briefly explained, on which the proposed method is based. Improvement details of the proposed method are presented in Section 4. Experimental results are shown and discussed in Section 5. Finally, the work is concluded in Section 6.

2. Related Works

There are many post-processing approaches to increase the resolution and accuracy of the ToF camera depth measurements. In general, the existing approaches can be divided into two categories: (1) fusion of a ToF sensor and a color sensor; (2) fusion with multiple depth sensor.

The first class of methods often fuse a ToF camera and a coupled color camera in a hybrid multi-modal camera rig. It is a convenient strategy to overcome the drawbacks of ToF camera by exploiting the advantages of each of the camera. The markov random field (MRF) model and simple linear iterative clustering are employed to generate high-quality depth map by optimizing global energy function [12]. Park *et al.*, [13] combine the MRF and the nonlocal means filtering to upsample the depth maps. However, they are computationally intensive and thus not suitable in real-time system. Joint bilateral filter and its variations [14, 15] are often used to increase the spatial resolution. Garica *et al.*, [14] proposed a novel unified multi-lateral filter. It can generate super resolution depth map and suppress the unwanted artifacts. Chan *et al.*, [15] proposed a noise-aware filter for depth upsampling. It preserves the benefits of using the joint bilateral filter while preventing the artifacts. The advantage of these bilateral filtering methods is they can be performed quickly, however they can often smooth fine details and cause edge blurring. Different from the aforementioned methods, Ferstl *et al.*, [16] formulated the depth image upsampling to a convex optimization by using the total generalized variation, but the edge inaccurate problem still exists.

The second class of methods often fuse multiple depth measurements in temporal domain into one depth map or fuse the ToF depth measurement with other depth sensor. Lee [17] proposed a novel temporal interpolative filtering method for ToF depth video. This method is not suitable for dynamic scenes. Stereo is often employed [18,19] to fuse with the ToF sensor. Liu *et al.*, [18] proposed a k-nearest neighborhood based integration method. Work by [19] proposed a reliable method by incorporating texture information, segmentation into a novel pseudo-two-layer model to improve the depth estimation. Work by [20] proposed a probabilistic method for fusing ToF data and stereo data based on mixed pixels measurement models. By using the complementary characteristics, these method show better results than the depth map by using the single sensor. However, the

passive stereo often fails within poorly textured regions, and these approaches are highly sensitive to the calibration effort.

3. TGV based Depth Upsampling Method [16]

This section briefly reviews the TGV based depth upsampling method in [16]. By using the combination of the low resolution ToF depth map D_L and the high resolution intensity image C_H, the high resolution dense depth map D_H is generated in a variational optimization framework by equation (1).

$$D_H = \arg\min_u \left\{ G(u, D_S) + TGV_\alpha^2(u) \right\} \tag{1}$$

Where $G(u, D_S)$ is the data term; $TGV_\alpha^2(u)$ is the second order TGV regularization term.

The data term is defined as equation (2).

$$G(u, D_S) = \int_{\Omega_H} w \left| (u - D_S) \right|^2 dx \tag{2}$$

Where $\Omega_H \subseteq R^2$, u is the reconstructed depth map in the iterative computational procedure. D_S is the high-resolution sparse depth image, which is mapped by the low-resolution depth image D_L according to the intrinsic and extrinsic parameters of the color sensor and the depth sensor. The value of the weighting factor w is set to $[0,1]$, which is zero at upmapped pixel points and between zero and one on the mapped pixel points.

The second order TGV regularization term is defined as equation (3). The anisotropic diffusion tensor $T^{\frac{1}{2}}$ is calculated by the equation (4)

$$TGV_\alpha^2(u) = \min_v \left\{ \alpha_1 \int_{\Omega_H} \left| T^{\frac{1}{2}} (\nabla u - v) \right| dx + \alpha_0 \int_{\Omega_H} \left| (\nabla v) \right| dx \right\} \tag{3}$$

$$T^{\frac{1}{2}} = \exp(-\beta \left| \nabla C_H \right|^\gamma) nn^T + n^\perp n^{\perp T} \tag{4}$$

Where α_1 and α_0 are the weighting factors. n is the normalized direction of the intensity image gradient, $n^\perp$ is the normal vector of the gradient. β and γ are set to 0.9 and 0.85, which are the adjust factors to balance the magnitude and the sharpness of the tensor.

The method is effective. However, it still has a few problems. The edge inaccurate problem still exists. The effect depends greatly on the weighting factors α_1 and α_0, and the two values are constant for each pixel and adjusted by hand for the different input images.

4. Proposed Method

Motivated by the work [16], the proposed method designed an improved TGV model to generate the high resolution dense depth map D_H by fusing the low-resolution depth image D_L with the high-resolution intensity image C_H。A diagram of the proposed method is presented in Figure 1.

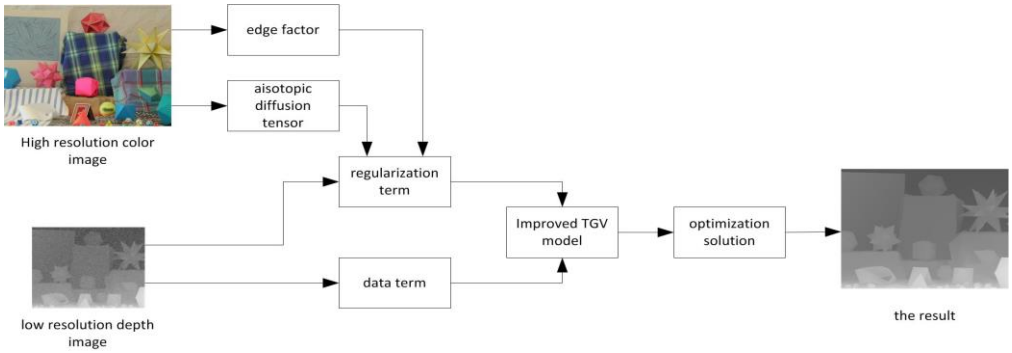


Figure 1. The Diagram of the Proposed Method

4.1. Improved TGV Model

In this paper, we design an improved TGV model. The improvements are reflected in two aspects. Firstly, an edge factor is introduced into the second order TGV regularization term. The edge factor is computed by the edge detection algorithm and given in Section 4.1.1. Secondly, the parameters adaptive mechanism about the regularization term is designed, which establishes the weighting factors α_1 and α_0 adaptively and updates them in the iteration procedure. The details are given in Section 4.1.2.

The improved TGV model is composed of the data term and the regularization term. The data term forces the result to be consistency to the input depth measurements. The data term $G(u, D_S)$ is defined in the equation (2). To obtain a high-resolution depth image by the low-resolution depth image is an ill-posed problem, the data term can constrain the deviation between the input depth image and the reconstructed depth image, but it cannot get the optimal solution. In order to get the high resolution dense depth map, the regularization term $TGV_\alpha^2(u)$ is defined by the internal relations between the intensity image and depth image of the same scene. The second order regularization term is sufficient for depth data regularization since most objects can be well approximated by piecewise affine surfaces. So the new regularization term $TGV_\alpha^2(u)$ is defined as the equation (5).

$$TGV_\alpha^2(u) = \min_v \left\{ \int_{\Omega_H} \alpha_1(x)\, g(x) \left| T^{\frac{1}{2}} (\nabla u - v) \right| dx + \int_{\Omega_H} \alpha_0(x) |(\nabla v)| dx \right\} \tag{5}$$

Where $\alpha_1(x)$ and $\alpha_0(x)$ are the adaptive parameters. $g(x)$ is the edge factor. The symbol x represents each pixel point. The details are given in the following parts of this section.

4.1.1. Edge Factor

Assuming that the edges in the intensity image usually correspond to the discontinuities of the depth map, the edge factor $g(x)$ is computed by the intensity image and it is introduced to the regularization term to handle the depth distortion.

Firstly, we compute the edge of the high-resolution intensity image C_H by using the multi-scale wavelet edge detection algorithm [21]. The three-order B-spline wavelet is adopted in this paper.

We represent the high-resolution intensity image C_H as $f(i, j)$. i and j are the pixel coordinates in horizontal and vertical directions.

By using 2-D smoothing function $\theta(i,j)$ to filter the image in different scales and calculate the first derivative, the wavelet transform has two components in the horizontal and vertical directions for any 2-D function $f(i,j)$.

$$\begin{bmatrix} WT_s^1 f(i,j) \\ WT_s^2 f(i,j) \end{bmatrix} = s \begin{bmatrix} \frac{\partial}{\partial i}[f(i,j) * \theta_s(i,j)] \\ \frac{\partial}{\partial j}[f(i,j) * \theta_s(i,j)] \end{bmatrix} \tag{6}$$

In equation (6), s is the scale. $WT_s^1 f(i,j)$ and $WT_s^2 f(i,j)$ respectively reflect the gradients in horizontal and vertical directions.

So the magnitude and the phase are given for each scale by the equation (7) and the equation (8).

$$M_s f(i,j) = \sqrt{\left|WT_s^1 f(i,j)\right|^2 + \left|WT_s^2 f(i,j)\right|^2} \tag{7}$$

$$A_s f(i,j) = \arctan\left[\frac{WT_s^2 f(i,j)}{WT_s^1 f(i,j)}\right] \tag{8}$$

We detect the local maximum value points of the magnitude of the wavelet along the gradient direction and the edge points of the image are obtained by adaptive threshold [21].

Secondly, we define the corresponding magnitude image as the equation (9). And we define the edge factor $g(x)$ as the equation (10).

$$m(x) = m(i,j) = \begin{cases} M_s f(i,j) & x \in \{edge\ point\} \\ 0 & others \end{cases} \tag{9}$$

$$g(x) = \frac{1}{1 + \rho \times m(x)} \tag{10}$$

Where the compensation factor ρ is designed and its value is set to 20 as an empirical value.

The value of $m(x)$ at the edge is larger than its corresponding value at the flat area. With the aid of the $g(x)$, the diffusion of the tensor can be suppressed effectively when the value of $g(x)$ is small at the edge. Then the details of the reconstructed depth map is better maintained, especially edges. By using the high resolution intensity data, we incorporate the gradient and edge information into the second order regularization term. The edge of the reconstructed high resolution dense depth map can be obtained accurately.

4.1.2. Adaptive Parameters

The weighting factors α_1 and α_0 are critical for the regularization term of the model. It is not appropriate by setting a constant for all the pixels. In order to improve the effect of the L1 minimization, a theoretical proof is presented for reweighting the factors in [22]. Based on the idea, we design a new parameters adaptive mechanism for the improved TGV model, where the weights are computed for each pixel and updated in each iteration. We compute the weighting factors as the equation (11).

$$\begin{cases} \alpha_1^{k+1}(x) = \dfrac{1}{|\nabla u^k(x) - v^k(x)| + \varepsilon} \\ \alpha_0^{k+1}(x) = \dfrac{1}{|\nabla v^k(x)| + \varepsilon} \end{cases} \tag{11}$$

Where k is the iteration number. ε is set to 0.1 in order to provide stability and the solution of the L1 minimization problem [22] is proved to be reasonably robust to the choice of ε.

4.2. Optimization Solution

According to the above equations (1),(2),(5),(10),(11), the objective function of the proposed method is defined as the equation (12) by the combination of the data term and the regularization term.

$$\min_{u,v} \left\{ \alpha_1 \int_{\Omega_H} g \left| T^{\frac{1}{2}}(\nabla u - v) \right| dx + \alpha_0 \int_{\Omega_H} |\nabla v| dx + \int_{\Omega_H} w |(u - D_S)|^2 dx \right\} \tag{12}$$

The primal-dual energy minimization scheme [23] is adopted to solve the optimization problem. By using the Legendre-Fenchel transform, the equation (12) can be reformed as the equation (13).

$$\min_{u \in R^{MN}, v \in R^{2MN}} \max_{p \in P, q \in Q} \left\{ \alpha_1 g \left\langle T^{\frac{1}{2}}(\nabla u - v), p \right\rangle + \alpha_0 \langle \nabla v, q \rangle + \sum_{i,j \in \Omega_H} w_{i,j} \left(u_{i,j} - D_{S_{i,j}} \right)^2 \right\} \tag{13}$$

Where p and q are the dual variables, and they are defined as the equation (14) and (15).

$$P = \left\{ p : \Omega_H \to R^2 \middle| \|p_{i,j}\| \le 1, \forall i, j \in \Omega_H \right\} \tag{14}$$

$$Q = \left\{ q : \Omega_H \to R^4 \middle| \|q_{i,j}\| \le 1, \forall i, j \in \Omega_H \right\} \tag{15}$$

The primal and dual variables are iteratively optimized for the individual pixels. For each iteration, the weighting factors $\alpha_1(x)$ and $\alpha_0(x)$ are updated by the equation (11). Then the variables are calculated by the equation (16).

$$\begin{cases} p^{k+1} = proj_p \left\{ p^k + \sigma_p \alpha_1 g (T^{\frac{1}{2}}(\nabla \bar{u}^k - \bar{v}^k)) \right\} \\ q^{k+1} = proj_q \left\{ q^k + \sigma_q \alpha_0 \nabla \bar{v}^k \right\} \\ u^{k+1} = \dfrac{u^k + \tau_u (\alpha_1 g \nabla^T (T^{\frac{1}{2}} p^{k+1}) + w D_S)}{1 + \tau_u w} \\ v^{k+1} = v^k + \tau_v (\alpha_0 \nabla^T (q^{k+1}) + \alpha_1 g T^{\frac{1}{2}} p^{k+1}) \\ \bar{u}^{k+1} = u^{k+1} + \lambda \left(u^{k+1} - \bar{u}^k \right) \\ \bar{v}^{k+1} = v^{k+1} + \lambda \left(v^{k+1} - \bar{v}^k \right) \end{cases} \tag{16}$$

Where $proj_p \left\{ \tilde{p}_{i,j} \right\} = \dfrac{\tilde{p}_{i,j}}{\max(1, |\tilde{p}_{i,j}|)}$, $proj_q \left\{ \tilde{q}_{i,j} \right\} = \dfrac{\tilde{q}_{i,j}}{\max(1, |\tilde{q}_{i,j}|)}$. $\sigma_p, \sigma_q, \tau_u, \tau_v$ are set to the default values in [23]. λ is updated in every iteration and computed as given in [23] in order to achieve a fast convergence.

4.3. The Step of the Proposed Method

The initial iteration values are set as $k=0, u^0=D_S, v^0, p^0, q^0=0$. The termination criterion for iteration is that the number of iteration is over the given maximum number or the difference of the variable u in adjacent iteration is less than the given threshold tol. In this paper, the maximum iteration number is set to 1000 and the threshold tol is set to 0.1. When the iteration is terminated, the final high resolution dense depth map D_H is obtained.

The main steps of the proposed method are presented in Table 1.

Table 1. The Main Steps of the Proposed Algorithm

Steps	Algorithm implementation
Step 1	**Input:** High resolution intensity image C_H and low resolution depth map D_L.
Step 2	Mapping D_L to high resolution sparse depth image D_S based on the result of the camera calibration process.
Step 3	**Parameters Initialization:** $k=0, u^0=D_S, v^0, p^0, q^0=0$ Computing the anisotropic diffusion tensor $T^{\frac{1}{2}}$ by the equation (4). Computing the edge factor $g(x)$ by the equation (10). Computer the weighting factors α_1^1, α_0^1 by the equation (11).
Step 4	Define the objective function as the equation (12). While (1) Compute and Update the variables by the equation (16); If $\left\|u^{k+1}-u^k\right\|<tol$ or k reach to the maximum limit, break; Else $k=k+1$. Computer the weighting factors α_1^k, α_0^k by the equation (11); end
Step 5	**Output:** the result D_H.

5. Experimental Results and Analysis

5.1. Setup

In order to test the proposed method, we have performed several quantitative and qualitative evaluations in this section. The public Middlebury datasets [24], which include the test images and the truth values, is adopted and the results are given in Section 5.2. Another test datasets of the real scene, which is provided in [25], is also adopted and the results are given in Section 5.3.

We have compared the state-of-the-art methods with the proposed method on the test datasets, including Bilinear interpolation algorithm, Ferstl's algorithm[16], Yang's algorithm[25], Park's algorithm[13]. These methods are implemented with the author's original codes, and these parameters are set to the default values.

The performance evaluation RMSE (Root Mean Square Error) is computed as in equation (17). The smaller the RMSE value, the better the effect of the reconstructed depth map.

$$RMSE = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M \times N} (x_i - y_i)^2} \tag{17}$$

Where $M \times N$ represents the number of pixels in rows and columns, x_i is the pixel value of the result of the specific depth map enhancement method, y_i is the corresponding true value.

5.2. Evaluation on the Middlebury Stereo Datasets

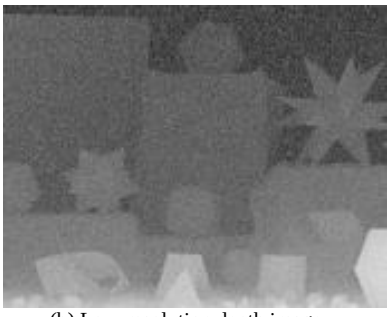
We choose the Middlebury Stereo Datasets to test the effectiveness of the method. Since the truth values are available, the quantitative and qualitative experiments are conducted.

The resolution of the original high-resolution color image and the depth map is 1390×1110. The low resolution depth map is generated by downsampling the original depth map. The downsampling factors is set to 2, 4, 6, 8 separately. The resolution of the results is the same as the resolution of the original color image.

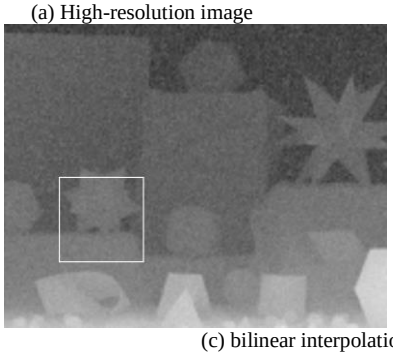
Two examples about the visual comparison for these methods are given in Figure 2 and Figure 3. The Moebius dataset is tested and results are presented in Figure 2. Figure 2(a) is the high resolution image, and Figure 2(b) is the low resolution depth map, the downsampling factor is 8. It is enlarged to facilitate the observation. Figure 2(c), 2(d), 2(e), 2(f), 2(g) are successively the results and the local enlarged images of the Bilinear interpolation algorithm, Yang's algorithm, Park's algorithm, Ferstl's algorithm, and our method. The local images are labeled with white rectangle in the corresponding source images.

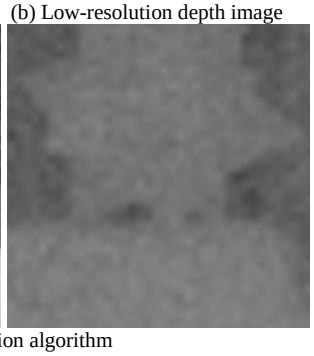
Also, the books dataset is tested and results are presented in Figure 3. Figure 3(a) is the high resolution image, and Figure 3(b) is the low resolution depth map, the downsampling factor is 8. It is enlarged to facilitate the observation. Figure 2(c), 2(d), 2(e), 2(f), 2(g) are successively the results and the local enlarged images of the Bilinear interpolation algorithm, Yang's algorithm, Park's algorithm, Ferstl's algorithm, and our method. The local images are labeled with white rectangle in the corresponding source images.

(a) High-resolution image

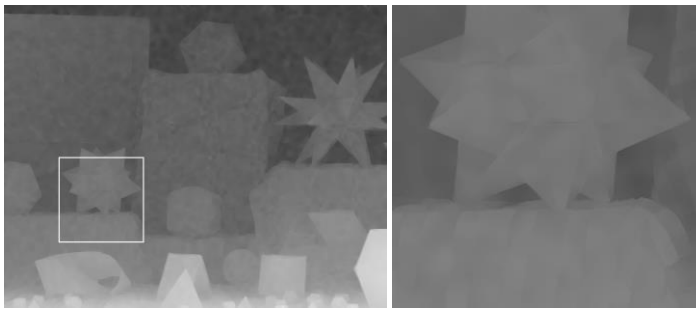


(b) Low-resolution depth image

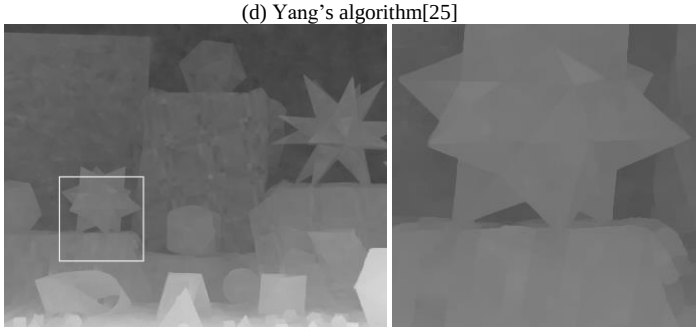




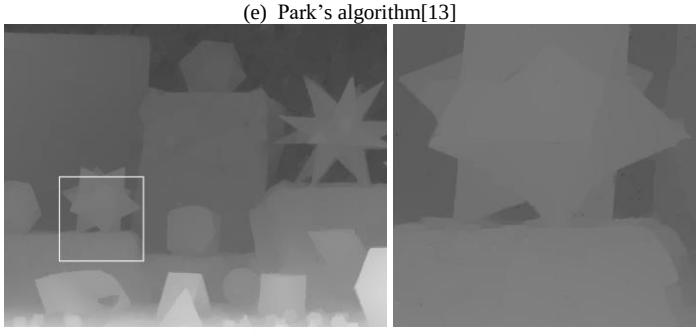
(c) bilinear interpolation algorithm



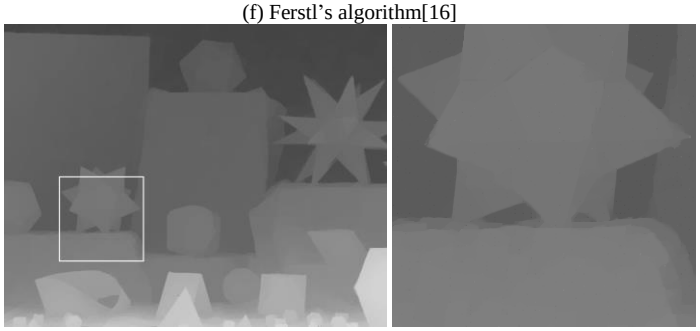
(d) Yang's algorithm[25]



(e) Park's algorithm[13]

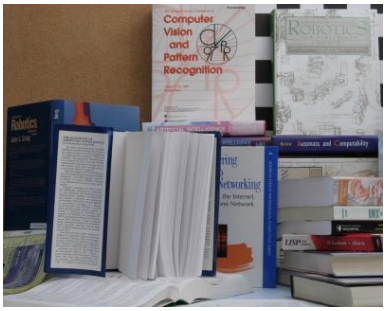


(f) Ferstl's algorithm[16]

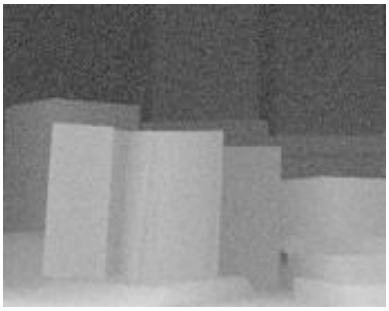


(g) our method

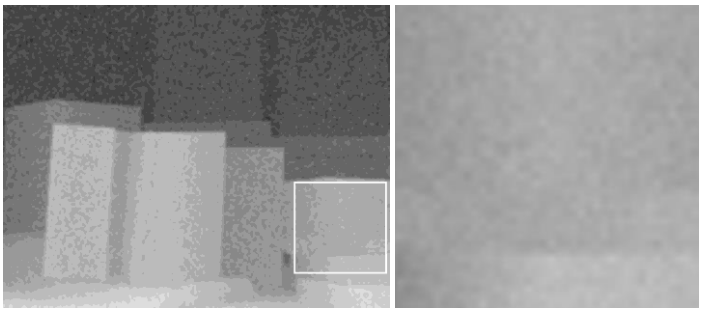
Figure 2. Visual Comparison of the Moebius Example in Middlebury Datasets



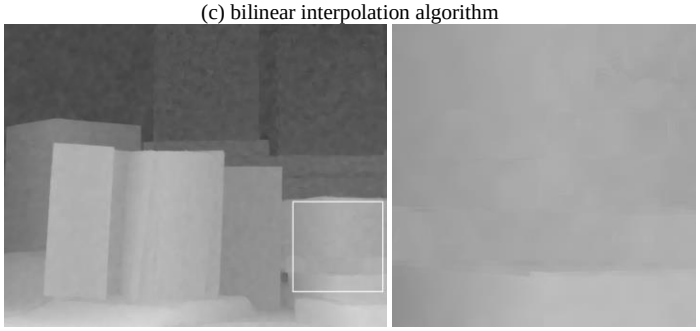
(a) High-resolution image



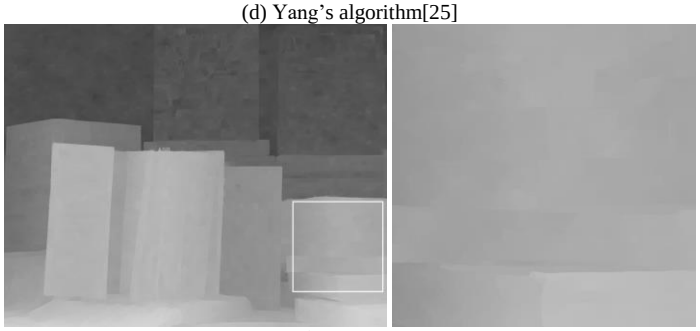
(b) Low-resolution depth image



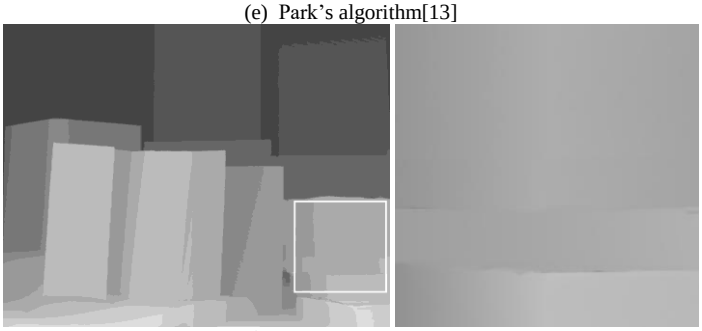
(c) bilinear interpolation algorithm



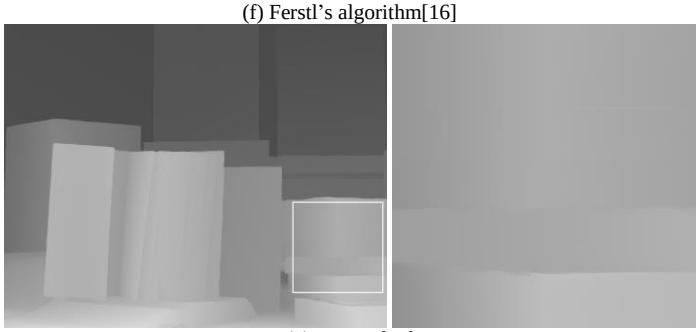
(d) Yang's algorithm[25]



(e) Park's algorithm[13]



(f) Ferstl's algorithm[16]



(g) our method

Figure 3. Visual Comparison of the Books Example in Middlebury Datasets

It is need to address that the high resolution image in the datasets is color image, so it is converted to the corresponding intensity image for evaluation. In Figure 2 and Figure 3 we present the color image for display.

As can be seen from the Figure 2, bilinear interpolation algorithm is less effective than the others, the results of Yang's algorithm and Park's algorithm are slightly blurred. Ferstl's algorithm can effectively suppress noise, but our method can not only suppress noise, but also highlight the edges, such as the contour of the objects is more clarity and accurate.

As can be seen from the Figure 3, the effect of the bilinear interpolation algorithm is the worst, the results of Yang's algorithm and Park's algorithm are slightly blurred. Ferstl's algorithm and our method can get better results, and more edges are preserved by our method, which can be seen in the edges of a stack of books.

The RMSE values with different downsampling factors (×2, ×4, ×8, ×16) are computed. The quantitative results are shown in Table 2. The RMSE values about Yang's algorithm, Park's algorithm are computed and given by [16].

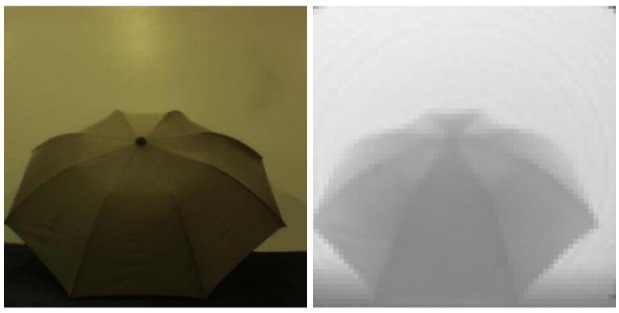
Table 2. Comparison of the RMSE

Method	Moebius				Books			
	x2	x4	x8	x16	x2	x4	x8	x16
Bilinear	4.20	4.56	4.87	5.43	3.95	4.31	4.71	5.38
Yang's algorithm[25]	1.92	2.42	2.98	4.40	1.87	2.38	2.88	4.27
Park's algorithm[13]	1.96	2.51	3.22	4.48	1.95	2.61	3.31	4.85
Ferstl's algorithm[16]	1.47	2.03	2.58	3.50	1.52	2.21	2.47	3.54
Our method	1.42	1.97	2.49	3.38	1.43	2.05	2.34	3.26

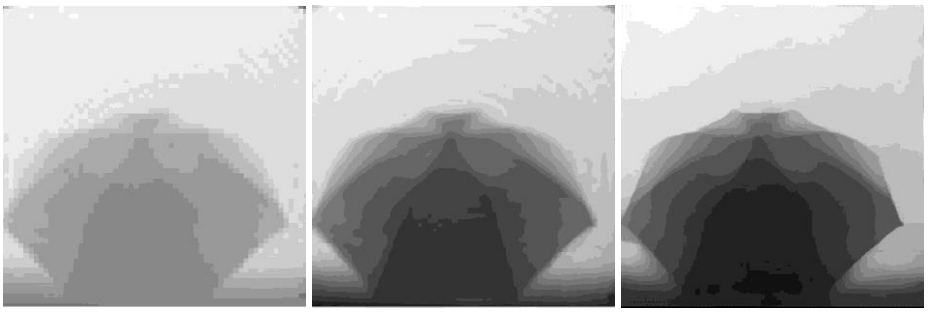
From the quantitative evaluation in Table 2, we can see that the RMSE value obtained by the proposed method is smaller than the other methods. It shows that the proposed method is more effective and it is consistent with the visual effect in Figure 2 and Figure 3. Also, it can be seen that the larger the downsampling factor, the larger the RMSE value. The phenomenon is consistent for each methods and it is easy to understand. Besides, the results of [16] and our result are superior to the others. It is need to addressed that the effect is depended on the parameters of α_1 and α_0 , which are fixed by hand in [16] and which are adaptive in our method.

5.3. Evaluation on the Datasets [25]

Another datasets are given in [25] and the images are captured by a color sensor and coupled ToF depth sensor. The FLEA digital color camera is used and its resolution is 640×640. The Canesta EP DevKit depth sensor is used to get the depth image and its resolution is 64×64. The resolution of the results is 640×640.

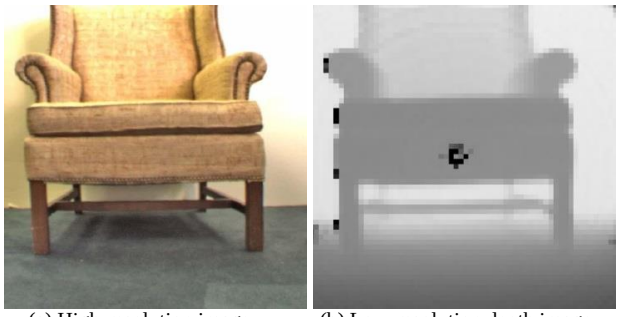


(a) High-resolution image (b) Low-resolution depth image

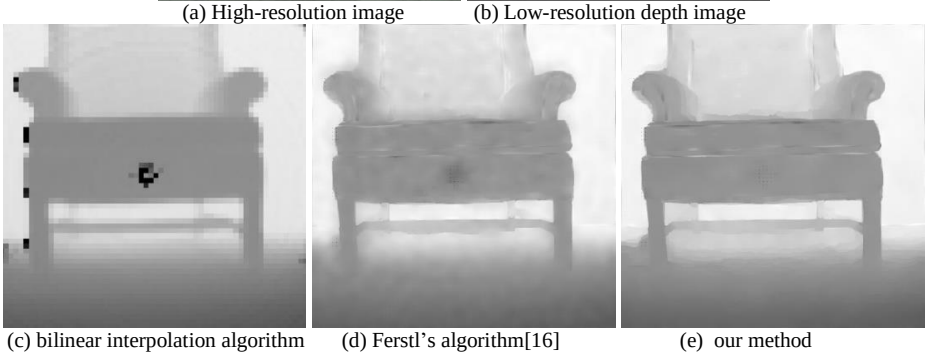


(c) bilinear interpolation algorithm (d) Ferstl's algorithm[16] (e) our method

Figure 4. Visual Comparison of the Umbrella Example



(a) High-resolution image (b) Low-resolution depth image



(c) bilinear interpolation algorithm (d) Ferstl's algorithm[16] (e) our method

Figure 5. Visual Comparison of the Chair Example

We only present the visual comparison results because there is no truth value about the datasets. Two examples about Umbrella and Chair are presented in Figure 4 and Figure 5. Figure 4(a) is the high resolution image, and Figure 4(b) is the depth image, Figure 4(c), 4(d), 4(e) are successively the results of the Bilinear interpolation algorithm, Ferstl's algorithm, and our method. Also, Figure 5(a) is the high resolution image, and Figure 5(b) is the depth image, Figure 5(c), 5(d), 5(e) are successively the results of the Bilinear interpolation algorithm, Ferstl's algorithm, and our method.

As it can be seen from Figure 4, the effect of the bilinear interpolation algorithm is the worst. The edge is inaccurate. The proposed method can obtain the high resolution depth image with accurate edges, as the edges of the umbrella in Figure 4(e) is more clear and smooth than the others.

The lack of the depth value is usually caused by the depth sensor. From the figure 5, we can see that the proposed method can cope with the holes in depth map and the consistency of depth value is better than the others.

5.4. Discussion

The experiments based on the two public datasets are conducted and the results are presented in Section 5.2 and 5.3. From the experiments, we can see that the proposed method can obtain a high resolution dense depth map with high quality. It can cope with

the common problems about low resolution, depth distortion, holes, and inaccurate edge. The reason is that the proposed method is designed and realized by incorporating the edge and gradient information of the high resolution intensity image into the Total Generalized Variation framework, and the weighting factors of the regularization term are adaptive for each pixel.

In this paper, the B-spline wavelet edge detection algorithm is adopted to realize the process of edge detection. Obviously, other state-of-the-art edge detection algorithm is easy to incorporate into the proposed method by replacing the edge factor $g(x)$. Besides, any type of the 2D high resolution mono/color image sensor with arbitrary resolution is applicable to the methods. The resolution of the depth map result depends on the resolution of the 2D image sensor. Another advantage of the method is the adaptive parameters of the regularization term, which can avoid adjusting the parameters manually and enhance the applicability in real scenes.

However, the computational complexity is higher than the others. The mainly time-consuming process is the optimization solution by iterations. It should be accelerated for real-time applications in the further study.

6. Conclusions

In the paper we propose a new depth map enhancement by fusing a ToF sensor and a coupled 2D high resolution color sensor. An improved total generalized variation model is presented to address the shortcomings of the ToF sensor such low resolution, noise, depth distortion. By using the accurate edge and gradient information of the high resolution intensity image and modifying the weighting factors adaptively, a high resolution and high quality depth map is obtained. We further provide the experimental results on the public datasets. The results show that the proposed method is effective and obtain better performance than the compared methods. As a future perspective, some accelerated strategies will be concerned for the real-time applications.

Acknowledgments

This Project is supported by the National Natural Science Foundation of China (Grant No. U1330133).

References

[1] S. Foix, G. Alenya and C. TorrasSmalley, “Lock-in Time-of-Flight(ToF) cameras: a survey”, IEEE Sensors Journal, vol. 11, no. 9, **(2011)**, pp. 1917-1926.
[2] J. Han, L. Shao, D. Xu, and J. Shotton, “Enhanced computer vision with Microsoft kinect sensor: a review”, IEEE Transactions on Cybernetics, vol. 43, no. 5, **(2013)**, pp. 1318-1334.
[3] A. A. Shetty, G. C. Nayak and V. I. George, “Stereo vision for mobile robots: a review”, International Journal of Engineering & Technology, vol. 2, no. 11, **(2015)**, pp. 25-29.
[4] R. S. Wang, “3D building modeling using images and LiDAR: a review”, International Journal of Image & Data Fusion, vol. 4, no. 4, **(2013)**, pp. 273-292.
[5] D. Piatti and F. Rinaudo, “SR-4000 and CamCube3.0 Time of Flight (ToF) Cameras: tests and comparison”, Remote Sensing, vol. 4, no. 4, **(2012)**, pp. 1-21.
[6] T. Tzschichholz, T. Boge and K. Schilling, “Relative pose estimation of satellites using PMD-/CCD-sensor data fusion”, Acta Astronautica, vol. 119, **(2015)**, pp. 25-33.
[7] L. Regoli, K. Ravndoor and M. Schmidt, “On-line robust pose estimation for rendezvous and docking in space using photonic mixer devices”, Acta Astronautica, vol. 96, **(2014)**, pp. 159-165.
[8] S. L. X. Francis, S. G. Anavatti, M. Garratt and H. Shim, “A ToF-camera as a 3D vision sensor for autonomous mobile robotics”, International Journal of Advanced Robotics Systems, vol. 12, no. 11, **(2015)**, pp. 1-15.
[9] S. Alenya, S. Foix and C. Torras, “ToF cameras for active vision in robotics”, Sensors and Actuators A: Physical, vol. 218, **(2014)**, pp. 10-22.
[10] Z. Zhang, “A flexible new technique for camera calibration”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, **(2000)**, pp. 1330-1334.

[11] J. Jung, J. Y. Lee, Y. Jeong, and I. S. Kweon, "Time-of-Flight sensor calibration for a color and depth camera pair", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 7, **(2015)**, pp. 1501-1513.
[12] G. T. Hao, X. P. Du, J. G. Zhao, H. Chen, J. J. Song and Y. S. Song, "Dense surface reconstruction based on the fusion of monocular vision and three-dimensional flash light detection and ranging", Optical Engineering, vol. 54, no. 7, **(2015)**, pp. 073113-1-07313-10.
[13] J. Park, H. Kim, Y. W. Tai, M. Brown and J. Kweon, "High quality depth map upsampling for 3D-TOF cameras", IEEE International Conference on Computer Vision, Barcelona, Spain, **(2011)** November 6-13.
[14] F. Garica, D. Aouada, B. Mirbach, T. Solignac and B. Ottersten, "Unified multi-lateral filter for real-time depth map enhancement", Image and Vision Computing, vol. 41, **(2015)**, pp. 26-41.
[15] D. Chan, H. Buisman, S. Theobalt and S. Thrun, "A noise-aware filter for real-time depth upsampling", Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, Marseille, France, **(2008)** October 18.
[16] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation", IEEE International Conference on Computer Vision, Sydney, Australia, **(2013)** December 1-8.
[17] S. Lee, "Time-of-flight depth camera accuracy enhancement", Optical Engineering, vol. 51, no. 8, **(2012)**, pp. 083203-1-083203-8.
[18] L. W. Liu, Y. Li, M. Zhang, L. H. Wang and D. X. Li, "K-nearest neighborhood based integration of time-of-fight cameras and passive stereo for high-accuracy depth maps", Journal of Zhejiang University-Science C (Computers & Electronics), vol. 15, no. 3, **(2014)**, pp. 174-186.
[19] J. Liu, C. P. Li, X. F. Fan and Z. Q. Wang, "Reliable fusion of stereo matching and depth sensor for high quality dense depth maps", Sensors, vol. 15, no. 8, **(2015)**, pp. 20894-20924.
[20] C. D. Mutto, P. Zanuttigh and G. M. Cortelazzo, "Probabilistic ToF and stereo data fusion based on mixed pixel measurement models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 11, **(2015)**, pp. 2260-2272.
[21] S. Q. Zhang, M. Hui, M. Liu, Z. Zhao, L. Q. Dong, X. H. Liu and Y. J. Zhao, "Edge detection based on adaptive threshold B-spline wavelet for optical sub-aperture measuring", International Conference on Optical Instruments and Technology, Beijing, China, **(2015)** May 17-19.
[22] E. J. Candes, M. B. Wakin and S. P. Boyd, "Enhancing sparsity by reweighted L1 minimization", Journal of Fourier Analysis and Applications, vol. 14, no. 5, **(2008)**, pp. 877-905.
[23] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging", Journal of Mathematical Imaging and Vision, vol. 40, no. 1, **(2011)**, pp. 120-145.
[24] D. Scharstein and C. Pal, "Learning conditional random fields for stereo", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, **(2007)** June 17-22.
[25] Q. X. Yang, R. G. Yang, J. Davis and D. Nister, "Spatial-depth super resolution for range images", IEEE Computer Vision and Pattern Recognition, Minneapolis, USA, (**2007**) June 17-22.

