https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-275

Multi-frame super-resolution utilizing spatially adaptive regularization for ToF camera

Haegeun Lee, Jonghyun Kim, Jaeduk Han, and Moon Gi Kang; School of Electrical and Electronic Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

Abstract

Recently, 3D time-of-flight cameras have been developed. The development enables utilization of depth images in various fields. However, acquired depth images are corrupted by noise during the image acquisition process and have relatively lower resolution than RGB images due to the limitation of ToF cameras. In this paper, a multi-frame super-resolution reconstruction algorithm is proposed for ToF depth images to overcome such limits. The purpose of the multi-frame super-resolution reconstruction is to reconstruct a high-resolution image from observed multiple low-resolution images through the sequential process of subpixel estimation and restoration. A conventional regularized super-resolution reconstruction algorithm which takes Tikhonov regularization has a major drawback of over-smoothing around edges. To overcome the disadvantage, the spatially adaptive regularization is suggested for preservation of edges. Experimental results show that the image reconstructed by the proposed super-resolution reconstruction algorithm contains significantly higher resolution with less amount of noise and sharper edges than the observed data.

Introduction

The advent of a time-of-flight (ToF) camera accelerates the growth of 3D imaging industries, such as 3D reconstruction, 3D TV, and robotics, by enabling us to acquire 3D depth information. 3D ToF cameras work by observing the reflected light after illuminating modulated light to the scene. The ToF cameras generally provide a depth image and an amplitude image. The former can be calculated by the phase shift between the source and the reflected light, and the latter is measured by the energy of reflected light respectively. The obtained depth image is usually utilized for 3D applications.

Often the acquired depth images often contain relatively lower resolution compared to color images due to the characteristics of ToF sensors, e.g. 176 x 144 for Mesa SR4000, 640x480 for Kinect [1]. In addition, the accuracy is quite limited as the receptor of a ToF is not exactly aligned with the light source. Images also can be contaminated with noise during the image acquisition process. The depth images with low resolution and low qualities limit the application in 3D reconstruction and delay the progress of 3D imaging.

To overcome these limitations, depth upsampling methods using high resolution color images have been consistently proposed. The methods utilize a color image, which is captured in the same scene of the corresponding depth image, to enhance the resolution of the depth image using filters such as joint bilateral filter [2], or guided filter [3]. However, as the color image cannot be exactly aligned with the depth image, and the color discontinuity and depth discontinuity cannot be the same, these methods often generate some artifacts.

Depth images also can be restored by applying multi-frame super-resolution reconstruction methods, which reconstruct a high-resolution image from observed multiple low-resolution images [4].

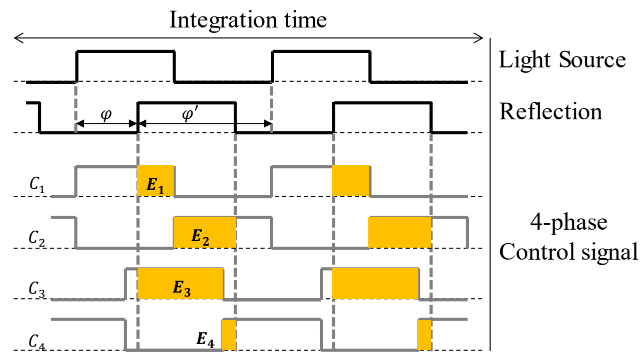

***Figure 1**. E_1 to E_4 indicate the electrical charges accumulated in the control signals C_1 to C_4.*

The multi-frame super resolution methods are employed in various applications, e.g. color images, satellite images, medical images, etc., so the methods also can be applied to depth images. The algorithms accompany the sequential process of subpixel estimation and restoration. To overcome the ill-posedness of the super-resolution reconstruction, many methods have been proposed. Hardie *et al.* formulated a Constrained Least Squares approach in the spatial domain and applied it to the multi-frame super-resolution reconstruction [5]. Lee and Kang proposed an improved version considering inaccurate subpixel registrations [6]. Patanavijit *et al.* proposed the Lorentzian-Tikhonov regularization to remove outliers [7]. However, as these algorithms were not devised for depth images, the results show relatively lower performance than that of other images, so the need for optimized algorithm is required.

In this paper, we propose the multi-frame super-resolution reconstruction algorithm which is suitable for depth images of ToF cameras. First, the subpixel registrations of low-resolution depth images are calculated by those of low-resolution amplitude images which are acquired at the same time of the corresponding depth images. Then edge information is measured by utilizing the weighted local variance [8] as a main tool for edge detection. Based on the above observations, we propose an optimization method that contains the spatially adaptive regularization which gives smaller weights in edge regions compared to flat regions in the smoothness term. As the proposed method preserves edge regions while smoothing flat regions, the algorithm is robust to regularization parameter. This allows the edges to be preserved while greatly reducing noise. Therefore, the high-resolution depth image with less noise can be restored by solving the problem.

The remainder of this paper is organized as follows. Section 2 gives the overall proposed algorithm with brief explanations of ToF depth measurement and multi-frame super-resolution reconstruction. In Section 3, The experimental results of the proposed method are demonstrated and compared to those of conventional methods. Finally, the conclusion of our paper is provided in Section 4.

Proposed Method

In this section, we give brief overviews of ToF depth measurement and the multi-frame super-resolution reconstruction, then explain the proposed method in detail.

ToF Depth Measurement

A time-of-flight camera utilizes infrared light source and the reflected light. An imaging sensor of the ToF camera receives the reflection and converts the photonic energy to electrical current. The light source emits modulated light, and the receptor detects only the reflection that is on the same spectrum of its control signals.

Figure 1 illustrates the operation of ToF sensing. The control signals demonstrate a phase difference of 90 degrees to each other, and the electrical charges induced by the reflection are measured during each of the signals. The phase shift between light and the reflection is calculated by the charges and can be translated to distance using the following formulation [1]:

$$\varphi = \tan^{-1}\left(\frac{E_3 - E_4}{E_1 - E_2}\right), \quad d = \frac{c}{4\pi f}\varphi, \tag{1}$$

where φ, d, c, and f represent the phase shift, distance, the speed of light, and the frequency of light respectively. $E_1 \sim E_4$ indicate the electrical charges accumulated in the corresponding control signals $C_1 \sim C_4$.

Amplitude also can be calculated by the energy of reflected light as below:

$$A = \frac{\sqrt{(E_1 - E_2)^2 + (E_3 - E_4)^2}}{2}. \tag{2}$$

The distance d in equation (1) and the amplitude A in equation (2) of each pixel can be combined into a depth image and an amplitude image respectively.

Multi-frame Super-resolution Reconstruction

The depth image degradation process contains warping followed by downsampling to get the observed low-resolution images from the high-resolution image. With an additive noise term in each low-resolution image, the observation model can be represented as

$$\mathbf{y}_k = \mathbf{D}\mathbf{M}_k\mathbf{x} + \mathbf{n}_k, \tag{3}$$

where the vectors $\mathbf{y}_k$, $\mathbf{x}$, and $\mathbf{n}_k$ denote the lexicographically ordered k th low-resolution depth image, the high-resolution depth image, and k th additive noise respectively. The matrix $\mathbf{M}_k$ and $\mathbf{D}$ are k th warping matrix and downsampling matrix respectively. The low-resolution depth image and the high-resolution image are denoted by $\mathbf{y}_k = \left[y_{k,1}, y_{k,2}, ..., y_{k,M}\right]^T$ for $k = 1, 2, ..., p$, and $\mathbf{x} = \left[x_1, x_2, ..., x_N\right]^T$, where $N = LM$ with the downsampling

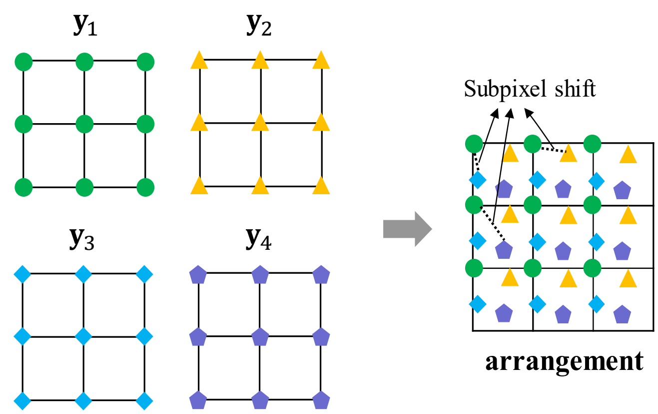

Figure 2. $\mathrm{y}_1 \sim \mathrm{y}_4$ are the images that have subpixel shift. The right side is arrangement of the images on the grid of y_1.

factor L. The matrix can be simplified, and the model is expressed as

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{n}_k, \tag{4}$$

where $\mathbf{H}_k$ is k th degradation matrix which includes warping and downsampling.

The purpose of the multi-frame super-resolution reconstruction is to obtain a high-resolution depth image $\mathbf{x}$ from observed multiple low-resolution depth images $\mathbf{y}_k$. The algorithm is the process of solving inverse problems in equation (3), finding $\mathbf{x}$ when $\mathbf{y}_k$ and $\mathbf{H}_k$ are given. While $\mathbf{y}_k$ can be observed by the ToF camera, $\mathbf{H}_k$ needs to be estimated. For this, the relative motion information between low-resolution images is necessary, and the reconstruction process consists of the sequential procedure of subpixel estimation and restoration.

Figure 2 indicates the arrangement of images that contain subpixel shift. The relative motions are measured through a subpixel estimation process. The subpixel estimation process requires abrupt spatial changes, especially sharp edges, and the changes can be represented as high frequency information in the frequency domain. Therefore, the lack of high frequency information in the depth images causes a problem in subpixel estimation. Instead of the depth images, the disadvantage is overcome by the utilization of amplitude images that contain enough high-frequency information for subpixel estimation. Therefore, the translation between observed low-resolution images can be calculated by usage of the amplitude images.

$$\begin{aligned}(\Delta i_k, \Delta j_k) &= \arg\min_{(\Delta i, \Delta j)} \left\| y_1(i,j) - y_k(i+\Delta i, j+\Delta j)\right\|^2 \\ &= \arg\min_{(\Delta i, \Delta j)} \left\| z_1(i,j) - z_k(i+\Delta i, j+\Delta j)\right\|^2,\end{aligned} \tag{5}$$

where $\mathbf{z}_1$ and $\mathbf{z}_k$ denote the amplitude images of the corresponding to depth images $\mathbf{y}_1$ and $\mathbf{y}_k$. (i,j) and $(\Delta i_k, \Delta j_k)$ mean the coordinates of the pixel of images and subpixel shift between $\mathbf{y}_1$

and $\mathbf{y}_k$ which is the same as that between $\mathbf{z}_1$ and $\mathbf{z}_k$. $\mathbf{H}_k$ can be derived from the subpixel shift acquired by equation (5)

Then for the restoration process that solves the inverse problem in equation (4), we employed regularization methods, which use prior knowledge to confine solutions in ill-posed problems. In this paper, we utilized the Tikhonov regularization method that employs high-frequency information for smoothness of an image [5]. The image acquisition is conducted by minimizing the functional

$$F(\mathbf{x}) = \sum_{k=1}^{p} \left\| \mathbf{y}_k - \mathbf{H}_k \mathbf{x} \right\|^2 + \lambda \left\| \mathbf{C}\mathbf{x} \right\|^2, \tag{6}$$

where $\mathbf{C}$ represents a high-pass filter, in this case Laplacian, and λ denotes the regularization parameter that controls tradeoff between fidelity to the data and smoothness of the image.

Spatially Adaptive Regularization

As the Tikhonov regularization in equation (6) applies the same smoothness over the depth image, ringing artifacts and over-smoothing problem appear on the edge regions of the reconstructed depth image. To overcome the drawbacks, we proposed spatially adaptive regularization, which employs different weights on edges and flat regions. For detection of edge regions, the weighted local variance is used as follows [8]:

$$\begin{aligned}
&\mathbf{q}_1 = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}^T, \\
&\mathbf{q}_3 = \begin{bmatrix} 0.25 & 0.5 & 0.25 \end{bmatrix}, \quad \mathbf{q}_4 = \begin{bmatrix} 0.25 & 0.5 & 0.25 \end{bmatrix}^T, \\
&\bar{\mathbf{x}}_l = \mathbf{q}_l * \mathbf{x}, \\
&\mathrm{var}_l(\mathbf{x}) = \mathbf{q}_l * (\mathbf{x} - \bar{\mathbf{x}}_l)^2,
\end{aligned} \tag{7}$$

where $\mathbf{q}_1 \sim \mathbf{q}_4$ are weight masks, and $\bar{\mathbf{x}}_l$ and $\mathrm{var}_l(\mathbf{x})$ are weighted local mean and weighted local variance corresponding to $\mathbf{q}_l$ for $l = 1,2,3,4$. The weighted local variance $\mathrm{var}_l(x_{i,j})$ shows relatively larger values in edge regions compared to flat regions or noise. The weighting functional using the weighted local variance is expressed as

$$\begin{aligned}
&\mathbf{W}_l = 1 - \frac{\log_{10}\left[1 + \mathrm{var}_l(\mathbf{x}) / T\right]}{\Theta_{\max}}, \\
&\Theta_{\max} = \max(\log_{10}\left[1 + \mathrm{var}_l(\mathbf{x}) / T\right]),
\end{aligned} \tag{8}$$

where T represents a threshold for the detection of edge regions, and $\Theta_{\max}$ is constant to normalize $\mathbf{W}_l$ from 0 to 1. By utilizing log function for weighting, the edge component of the depth image has a significantly smaller value of weight than flat regions or noise. Figure 3 displays the visualization of $\mathbf{W}_l$.

The minimization functional for reconstruction is as follows:

$$F(\mathbf{x}) = \sum_{k=1}^{p} \left\| \mathbf{y}_k - \mathbf{H}_k \mathbf{x} \right\|^2 + \lambda \sum_{l=1}^{4} \left\| \mathbf{x} - \mathbf{Q}_l \mathbf{x} \right\|_{\mathbf{W}_l}^2, \tag{9}$$

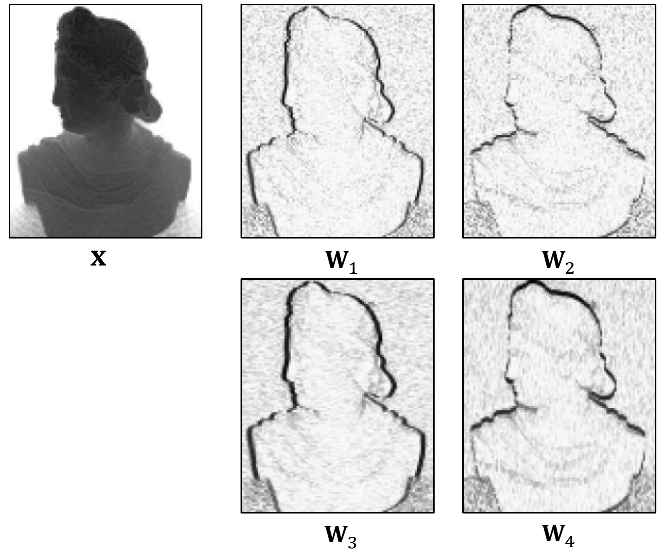

***Figure 3**. $\mathbf{x}$ is the depth image for edge detection, and $\mathbf{w}_1$ to $\mathbf{w}_4$ are the images of weights that are calculated by equation (7) corresponding to weight masks $\mathbf{q}_1$ to $\mathbf{q}_4$.*

where $\mathbf{Q}_l$ is matrix form of the weight masks $\mathbf{q}_l$. The weighting functional $\mathbf{W}_l$ corresponds to the direction of edges in regularization. Therefore, even if large regularization parameters are applied to noisy images, the edges can be preserved while the noise is significantly reduced.

Experimental results

The proposed method is investigated with degraded depth images of two different objects. The depth images are captured by ToF cameras with size of 224×172. As the captured depth images contain low resolution and low quality to be downsampled, the experiment was conducted by multiple captures of the same object with subpixel registration. Therefore, a reference image for evaluation of performance cannot exist, so we verified the performance by using subjective tests. We evaluated the quality of the images according to the distortion factors like discontinuities, noise, and artifacts. 9 depth images were utilized for the 900% resolution enhancement experiment.

Figure 4 illustrates the experimental result images. Figure 4 (a), (e) are captured images by ToF cameras, which are degraded by the cameras and environments. Discontinuities on edges and a large amount of noise are observed. Figure 4 (b), (f) are interpolated images with conventional method, specifically bicubic interpolation. The discontinuities are improved compared to Fig. 4 (a) and (e), but the noise still exists. Figure 4 (c), (g) are reconstructed images with multi-frame super-resolution reconstruction with Tikhonov-regularization, and both the discontinuities and noise are reduced. However, the edge regions in Fig. 4 (c) are over-smoothed, and ringing artifacts are generated outside of the edges in Fig. 4 (g). Figure 4 (d), (h) are restored images with prop osed method, which contains sharp edges and less ringing artifacts with reduced noise.

Conclusion

In this paper, we proposed a multi-frame super-resolution reconstruction method utilizing spatially adaptive regularization for ToF cameras to restore low-resolution degraded images to the high-

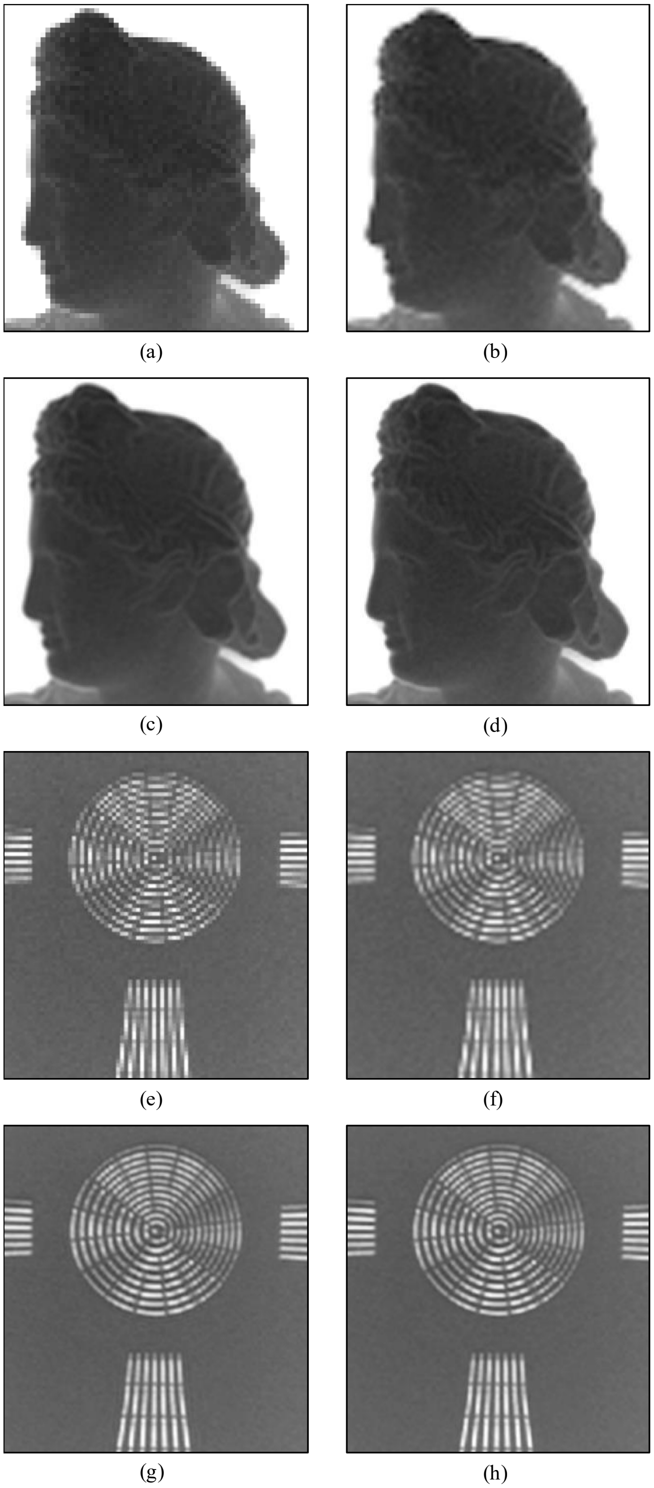

Figure 4. Magnified experimental results: (a),(e) captured image, (b),(f) single image interpolation, (c),(g) multi-frame super-resolution reconstruction with Tikhonov-regularization, (d),(h) proposed method.

resolution high-quality depth image. By applying a multi-frame super-resolution reconstruction method, the proposed method enables usage of abundant sources of the depth images and suggests another way to depth upsampling.

Subpixel estimation process is generally conducted by comparing the pixel information of the image features and details. Therefore, the utilization of depth images for the materials of the subpixel estimation is hard, because the depth images have deficient high-frequency information to measure the subpixel. We overcame the disadvantage by utilizing amplitude images that contain quite more details than depth images. Furthermore, the high-resolution depth images with less noise and high qualities are reconstructed by multi-frame super-resolution reconstruction via spatially adaptive regularization that employs weighted local variance and log function. The experimental results show the improved quality of the restored image compared to conventional L2-regularized method, specifically sharp edges and less ringing artifacts.

The proposed method is expected to be used to restore degraded depth images for ToF camera, and the restored depth images would enable the acquisition of high-quality 3D images.

References

[1] M. Hansard, S. Lee, O. Choi, and R. Horaud, Time-of-Flight cameras: Principles, Methods and Applications, Springer, 2012.

[2] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint Bilateral Upsampling," ACM Trans. Graph., vol. 26, no. 3, 2007.

[3] K. He, J. Sun, and X. Tang, "Guided Image Filtering," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, pp. 1397-1409, 2013.

[4] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," IEEE Signal Process. Mag., vol. 20, no. 3, pp. 21-36, 2003.

[5] R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E. Armstrong, and E. A. Watson, "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," Opt. Eng., vol. 37, pp. 37, 1998.

[6] E. S. Lee and M. G. Kang, "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," IEEE Trans. Image Process., vol. 12, no. 7, pp. 826-837, 2003.

[7] V. Patanavijit and S. Jitapunkul, "A Lorentzian Stochastic Estimation for a Robust Iterative Multiframe Super-resolution Reconstruction with Lorentzian-Tikhonov Regularization", ACM EURASIP J. Adv. Signal Process, vol. 2007, no. 2, pp. 21, 2007.

[8] S. C. Park and M. G. Kang, "Noise-adaptive edge-preserving image restoration algorithm," Opt. Eng., vol. 39, pp. 39, 2000.

Author Biography

Haegeun Lee received his BS degree in electrical and electronic engineering from Yonsei University, South Korea, in 2016. He is currently pursuing a Ph.D. degree with the School of Electrical and Electronic Engineering at Yonsei University. His current research interests include superresolution image processing and image restoration.

Jonghyun Kim received his BS degree in electrical and electronic engineering from Yonsei University, South Korea, in 2012. He is currently pursuing a Ph.D. degree with the School of Electrical and Electronic Engineering at Yonsei University. His current research interests include superresolution image processing, color interpolation and image restoration.

Jaeduk Han received his BS. degree in electronics engineering in 2014 from Yonsei University, Seoul, South Korea. He is currently pursuing a Ph.D. at the Department of Electrical and Electronic Engineering at Yonsei

University. His current research interests include the inverse problem in image and video processing.

Moon Gi Kang received Ph.D. degree in electrical engineering from Northwestern University, USA in 1994. He is currently a professor at the Department of Electronic Engineering of Yonsei University, Korea. He has served as the Editor of SPIE Milestone Series Volume, the Guest Editor of the IEEE SP Magazine Special Issue on Superresolution Image Reconstruction (May 2003), the Associate Editors of the EURASIP Journal of ASP and the Elsevier Journal of Digital Signal Processing.