2010 International Conference on Cyberworlds

Fast and simple super resolution for range data

Xueqin Xiang, Guangxia Li, Jing Tong, Zhigeng Pan
State Key Lab of Computer Aided Design and Computer Graphics
Zhejiang University
Hangzhou, China
{xiangxueqin,lgx,tongjing,zgpan }@cad.zju.edu.cn

***Abstract*—Current active 3D range sensors, such as time-of-flight cameras, enable acquiring of range maps at video frame rate. Unfortunately, the resolution of the range maps is quite limited and the captured data are typically contaminated by noise. We therefore present a simple pipeline to enhance the quality as well as improve the spatial and depth resolution of range data in real time by upsampling the depth information with the data from high resolution video camera and utilizing a new strategy to increase the sub-pixel accuracy. Our algorithm can greatly improve the reconstruction quality, boost the resolution of the range data to that of video sensor while achieving high computational efficiency for a real-time application.**

Keywords-super resolution; fast multi-lateral filter; sub-pixel estimation.

I. INTRODUCTION

In recent years, a variety of range measureing devices have been developed for 3D data acquisition. For example, by using extremely faster shutter (on the order of nanosecond), Time of Flight(TOF) sensors [1]measure time delay between transmission of a light pulse and detection of the reflected signals on an entire frame at once which are best suited for dynamic scene. The main contender to TOF sensor- stereo vision [3]- is rather limited: it is known to be quite fragile in practice (e.g. due to lack of texture).

Unfortunately, being a relatively young technology, TOF sensors have not enjoyed the same advances, with respect to image resolution, quality and photo speed, that have been made in traditional 2D intensity imaging sensors. As a result, in current generation, these sensors provide noise-contaminated range data of comparably low image resolution (e.g. only up to 176 × 144 for MESA Swissranger™ SR3000 [2]).

In this paper, we therefore propose a simple framework to substantially enhance the spatial and depth resolution of range data, e.g., those from the Mesa imaging sensor. To achieve this goal, firstly, we propose a new fast multi-lateral filter, termed FMLF, to adaptively upsample the low resolution range data in real time by taking advantage of the significant information provided by registered high resolution video camera. Secondly, to enhance the depth resolution and reduce the discontinuities caused by quantization in the depth map initiation process, a sub-pixel estimation algorithm then is formulated as a Markov Random Field (MRF) and treated it as a Maximum A Posteriori (MAP) problem.

The main contribution of our method is to present a simple pipeline to enhance the spatial and depth resolution of range data while obtaining real time performance. We also extend our method into a new realm: combined with the low resolution intensity image generated by TOF sensor itself, the quality of range data can be greatly improved.

The rest of the paper is organized as follows. Section Ⅱ introduces the related works. The complete description of the proposed fast and simple super resolution technique is presented in Section Ⅲ. The extension is given in Section Ⅳ. The experimental results are given in Section Ⅴ. Finally the conclusions are outlined in Section Ⅵ.

II. RELATED WORKS

There are many approaches that exploit additional information to improve the resolution of range data combining TOF sensor with one or two high resolution video cameras. Some of these techniques use a probabilistic approach: In [5], MRF is first designed based on the low resolution depth maps and the high resolution camera images. Unfortunately, this method gives promising spatial resolution enhancement only up to 10×. Yang et al. [6] then present a method modals a cost volume of depth probability and iteratively applies bilateral filter [7] to refine the cost volume. However, they do not use a joint bilateral filter [8] to link the two images and even with GPU (Graphics Processing Unit) [9] optimization, their effective runtime would be very large due to the number of cost slices and the iterative scheme. Another recent method [10] utilizes exclusively depth maps, without color image aid: a sequence of low resolution depth maps of same scene is aligned and then merged together to obtain a single depth map with improved resolution. But this method is restricted to static scenes' acquisition.

Key to our success is the use of multi-lateral filter, which essentially is the extension of joint bilateral filter widely used in several state-of-the-art upsampling algorithms [11]. Until recently, these edge-preserving bilateral filters were too computationally intensive for real time applications. Several efficient metho-

978-0-7695-4215-7/10 $26.00 © 2010 IEEE
DOI 10.1109/CW.2010.58

ds [12] enable it to be computed at constant time or even video using GPU implementation [13]. Yang et al. [14] improve on this by not explicitly representing the entire space, but instead sweeping a plane through the intensity level, computing the output in intensity order. This low-memory, cache-friendly algorithm is the fastest known bilateral filter. Inspired by Yang's acceleration strategy, our multi-lateral filter is sliced into one bilateral filter and one joint bilateral filter that computed through discretization technology respectively. Therefore, the real-time performance can be eventually achieved via GPU implementation. What's more, compared with the work [16], our method not only allows *arbitrary* function in filter, but also considers sub-pixel accuracy, in contrast with a potentially blocky range result.

TOF sensor also provides an intensity image that is perfectly registered with a depth map at each frame. Since a little interest has been put into this realm [15], we extend our algorithm for improving the quality of range data of TOF sensor by its own low resolution intensity image.

III. ALGORITHM

An overview of the framework of the approach is given out in Figure1 and it has two main independent phases: First, up-sample the low resolution range image from TOF sensor to the same size as the high resolution camera image and fast multi-lateral filter (FMLF) is applied for the purpose of spatial super resolution and denoising afterwards. In contrast to Chan's method [16], our fast multi-lateral filter enables of *arbitrary* spatial function and *arbitrary* range function. To reduce the quantization effect of the depth map (i.e. for the enhancement of depth super resolution), then a sub-pixel refinement algorithm is proposed based on probabilistic model. We will explain the details below.

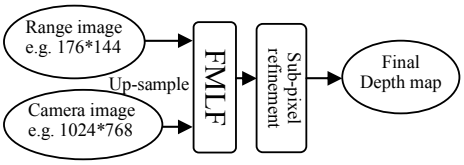


Figure 1. Pipeline of our algorithm. The range image is up-sampled to the same size as the camera image, and two different images serves as the inputs of the fast multi-bilateral filter. The following is a sub-pixel refinement process.

A. FMLF for depth upsampling

To cope with the spatial super resolution requirement and meanwhile denosing for noisy real-time 3D sensors, like time-of-flight cameras, we propose a new fast multi-lateral filter for upsampling (FMLF). It is our goal to satisfy spatial super resolution and denosing requirement for real-time 3D sensors as fast as possible and to make our filter to be more flexible. Like [16], the FMLF filter takes the following form:

$$I_x^F = \frac{1}{k_p}\sum_{y\in N(x)} I(y) f_S(x,y)[(1-\lambda(\sigma^2)) f_{\tilde{R}}(D(x),D(y)) + \lambda(\sigma^2) f_R(I(x),I(y))] \quad \text{----(1)}$$

Where x is a pixel in low resolution range image and y is a pixel in the neighborhood of $N(x)$, $I(x)$ and $I(y)$ are the corresponding range values of pixel x and y, $D(x)$ and $D(y)$ denote the intensity values of pixel x and y in high resolution camera image respectively, f_S, $f_{\tilde{R}}$ and f_R are all *arbitrary* functions, e.g. Gaussian function or Box function, $\lambda(\sigma^2)$ represents a blend function, defining in the interval [0,1], σ^2 is the variance in pixel neighborhood

From the equation (1), it is easy to conclude that a low weight λ makes our filter behave like a standard joint bilateral filter while a high weight λ gives higher influence to the latter range term $f_R(I(x),I(y))$ which makes the filter behave like an edge preserving bilateral filter that smooths the 3D geometry independently of information from the high resolution camera image. The crux is to decide the value of weight λ since it determines the characteristic of our filter. We want the filter to switch to a bilateral filter in cases where the areas are actually smooth but heavily contaminated with random noise caused by range measure. Therefore, we intuitively define our blend function $\lambda(\sigma^2)$, which does not need a preprocessing [16], as follows:

$$\lambda(\sigma^2) = \frac{\tau}{\sigma^2+\tau} \quad (2)$$

Here, σ^2 is the variance in pixel neighborhood *N*. Once σ^2 being low, latter range term $f_R(I(x),I(y))$ will be triggered and our filter will act as a bilateral filter to ease the errors caused by range measurement. The unique parameter τ depends on the characteristic of the employed depth sensor and can be determined through experiments.

The complexity of Eq. (1) makes direct compute could be time consuming and it is infeasible for real-time application. Several efficient numerical schemes [14] [17] have been proposed to reduce the computational load of bilateral filter. Inspired by the fastest bilateral filter method [14] so far, our filter is sliced into one bilateral filter and one joint bilateral filter as follow:

$$I_x^F = \frac{1}{k_p}\sum_{y\in N(x)} (1-\lambda(\sigma^2)) f_S(x,y) f_{\tilde{R}}(D(x),D(y)) I(y) + \frac{1}{k_p}\sum_{y\in N(x)} \lambda(\sigma^2) f_S(x,y) f_R(I(x),I(y)) I(y)$$

----(3)

Here, the former is a joint bilateral filter while the latter is a bilateral filter. We then could take advantage of acceleration technology proposed by [14]: the range data of low resolution range image and the intensity data of high resolution camera image are discretized into a number of values, compute a linear filter for each such value respectively, the output of which is termed as PBFIC in [14] and get intermediate results by a linear interpolation between two closest PBFICs. The final result is obtained through adding operation between intermediate results. Owing to the acceleration strategy discussed above, our GPU implementation of FMLF runs at about 35 frames per second using 8 PBFICs.

B. Sub-pixel estimation

We obtain disparities of the range image on integer level after the process detailed in section above. There has been a growing interest [18] in obtaining accurate sub-pixel disparity since the parabola fitting approaches exhibit artifacts known as pixel-blocking [19]. With the help of Fourier analysis, Scharstein and Szeliski [20] has concluded that sinc interpolator is in theory the best interpolation to evaluate the disparity space image at fractional disparities. Assume that the Markov Random Field (MRF) is represented by a set Φ of potential functions over cliques in an undirected graph $G = (V,D)$, where V represents a random variable set and D is the edge set, which includes the dependencies between every two variables in V. Our approach treats the sub-pixel estimation as energy minimization problem with:

$$E_{tot} = \sum_{p \in V} E_p(d_p) + \sum_{(p,q) \in D} \alpha E_s(d_p, d_q) \quad (4)$$

where data term E_p is the cost of assigning disparity d_p to pixel p, pairwise smoothness term E_s is the cost of assigning labels d_p and d_q to two neighboring pixels and α is a scale factor. The higher that α is chosen, the smoother is the resulting disparity map.

Let d_{int} be the integer disparity computed by our fast multi-lateral filter. The data cost of choosing d_p unequal to the former estimated d_{int} is formulated as following:

$$E_p(d_p) = (d_p - d_{\text{int}})^2 \quad (5)$$

Let $\tilde{d}$ be the average disparity within the considered patch D. The smoothness term E_s is defined according to:

$$E_s(d_p, d_q) = (d_p - \tilde{d})^2 \quad (6)$$

Since our energy Eq. (4) has a simple form, it is easy to compute the best solution for a certain point directly instead of to inference by belief propagation (BP)[4].Partial derivation $\partial E_{tot} / \partial d_p = 0$ yields

$$d_p = \frac{d_{\text{int}} + \alpha(N-1)/N * \tilde{d}}{1 + \alpha(N-1)/N} \approx \frac{d_{\text{int}} + \alpha * \tilde{d}}{1+\alpha} \quad (7)$$

Where N is the number of pixels within considered patch D.

In order to get close to the best solution of the above described problem, we need to iterate Eq. (7) to propagate the update disparity values: d_{int} remains the origin value while d_p is updated in every iteration. See Section Ⅴ for results of our sub-pixel estimation.

IV. EXTENDED RANGE DATA SUPER RESOLUTION BASED ON A SINGLE TOF SENSOR

TOF camera robustly provides a range image of real world scenes at video frame rates that is perfectly registered with an intensity image. At this point, it looks like an ordinary color camera plus additional range information. We extend our range data super resolution *only* with a single TOF sensor, based on the insight that range measurement may be improved according to the low resolution grayscale intensity image of TOF sensor itself.

Unlike [10], our method relies on one frame and it does not require the setup or calibration process as literature [21] did previously. Therefore, it is available for real-time application, especially within dynamic environment.

Assume the $\tilde{I}$ denote the low resolution grayscale intensity image obtained from TOF sensor. The fast multi-lateral filter we used is changed into following:

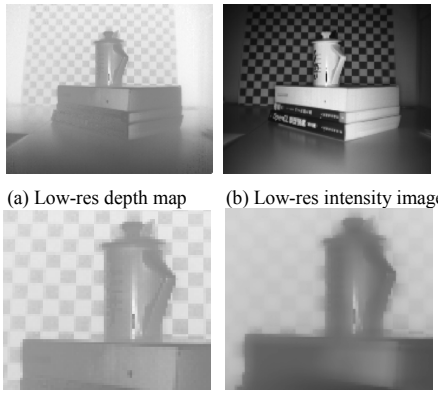
$$I_x^F = \frac{1}{k_p} \sum_{y \in N(x)} I(y) f_S(x,y) [(1-\lambda(\sigma^2)) f_{\tilde{R}}(\tilde{I}(x), \tilde{I}(y)) + \lambda(\sigma^2) f_R(I(x), I(y))]$$

----(8)

This is almost identical to Eq. (1) with the exceptions that the high resolution camera image is substituted with the low resolution grayscale intensity image. The sub-pixel refinement strategy detailed in section Ⅲ is also utilized to reduce quantization effect.

As shown in Fig. 2, our method successfully improves the quality of the low resolution depth maps. True geometry detailed in data, such as discontinuities, are preserved and enhanced, the random noise level is greatly reduced. Note that the generic artifacts that arise from the sensitivity of TOF sensor to object reflectance [21] are also prevented. By exploiting the GPU as a fast stream processor,

real time performance is feasible. In a word, our design successfully handles the data produced by state-of-the-art time-of-flight sensors which exhibit significantly higher random noise levels than most active scanning devices.



(a) Low-res depth map (b) Low-res intensity image

(c) Raw depth map(zoomed) (d) Refined result (zoomed)

Figure 2. From a low resolution depth map (a) and a low resolution grayscale intensity map (b) we create a depth map at a higher level quality. The significantly higher quality of our refined result (d) as opposed to the raw depth (c) is obvious.

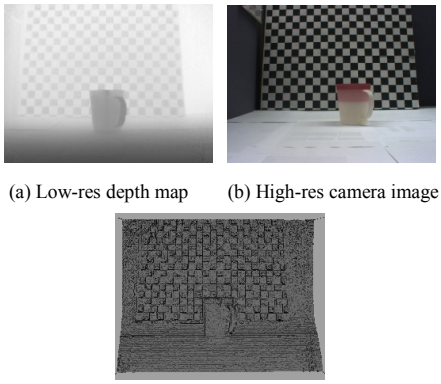
V. EXPERIMENTS

Our experimental system consists of a Mesa Swissranger^{TM} SR3000 time-of-flight camera and a Point Grey^{TM} dragonfly2 video camera. The two cameras are placed side-by-side (as closely as possible) and are frame-synchronized. The Swissranger can produce range images with size up to 176 × 144 pixels and the dragonfly2 can provide color images with resolution up to 1024 × 768. To align the range and video images, we resort to the gray-scale intensity images that the Swissranger sensor provides in addition to range images. Given this, the approach [22] is applied for image registration.

Compared with previous work [16] [21], only two main parameters are involved in our algorithm, they are τ and α . τ is the constant used in Eq.(2) which essentially denotes the expected variance due to noise, it is set to 50 experimentally. α is the magnification ratio of smooth term in Eq.(4) and is set to 2 in this paper.

A. spatial super resolution

Let the f_S, $f_{\tilde{R}}$ and f_R in Eq. (1) be all Gaussian functions for the purpose of a fair comparison, we evaluate our algorithms on a real scene where a checkerboard and a cup are involved as shown in Fig. 3. It is clear that our method successfully upsamples the low resolution depth maps to high resolution and with respect to the raw 3D data, the visual appearance of depth detail of the checkerboard is improved, especially on textured regions and around boundaries. Our approach is also superior to the Chan's method proposed by [16]. Please pay attention to the details, indicated by the red boxes and arrows, for further comparisons. Furthermore, evaluated on Nvidia Geforce 9800 GT platform, the GPU implementation of our approach averages 31ms which is faster than that of the Chan's method (37 ms).



(a) Low-res depth map (b) High-res camera image

(c) Raw 3D data

(d) Refined 3D data using Chan's approach

(e) Refined 3D data using our approach

Figure 3. By using of the high-res camera image (b), our technique upsamples a low-res depth map (a) reconstructed as 3D geometry(c) to a high-res depth map which can be reconstructed as 3D geometry (e) with the comparison of Chan's results (d).

A visual comparison of the depth maps of the middlebury datasets are provided in Fig. 4. The original depth map is down-sample by 8 (2^3) from the ground truth. Currently, f_S is chosen to be Box function, $f_{\tilde{R}}$ and f_R in Eq. (1) are all chosen to be Gaussian functions. Cleary, the results using our approach have more clean edges than the input depth maps and the result using MRF approach [5]. According to fig. 4, it is faithfully

acknowledged that our results are inferior to the results using Yang's method [6]. However, our approach is designed based on fast and simple pipeline whereas the Yang's method relies on iterations which make it impossible for real-time application.

Fig. 5 show the performance of our algorithm on middlebury datasets when f_S is chosen to be Box function or Gaussian function. Clearly, the two curves are almost coincidence in this experiment. However, we have found that Gaussian function is more robust, especially for noisy cases.

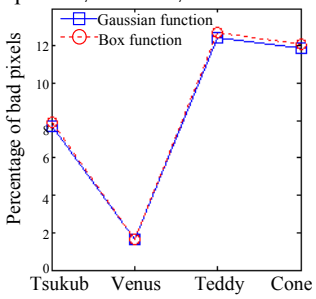
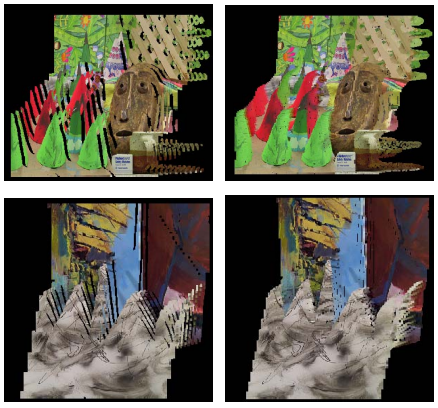


Figure 5. The performance of our algorithm on Middlebury datasets with regard to f_S being Box or Gaussian function (with error threshold 1).

B. depth super resolution

Besides the enhancement of the spatial resolution of range images, our approach also provides sub-pixel estimation for the further enhancement of the depth resolution of range images. A set of synthesized views are shown is Fig. 6, providing a visual comparison of the algorithms with and without sub-pixel refinement. The enhancement of the depth resolution is clear: As shown in column (a), the results are quantized into discrete number of planes. After sub-pixel estimation, the quantization effect is removed, as it is shown in column (b).



(a) Before (b) After

Figure 6. Synthesized views produced by our approach before or after sub-pixel estimation.

Table 1 evaluates the performance of our approach with and without sub-pixel estimation on Middlebury datasets. The original depth map is down-sample by 8 (2^3) from the ground truth. By comparing bad pixel percentages with and without sub-pixel estimation, we can conclude that sub-pixel refinement improves the performance of our approach for all data sets.

TABLE I. COMPARISON OF THE RESULTS ON MIDDLEBURY DATASETS WITH OR WITHOUT SUB-PIXEL REFINEMENT (WITH ERROR THRESHOLD 1)

	Tsukuba	*Venus*	*Teddy*	*Cone*
Without sub-pixel refinement	8.23%	1.73%	13.5%	12.9%
With sub-pixel refinement	7.71%	1.62%	12.4%	11.9%

VI. CONCLUSIONS

In this paper, we present a fast and simple framework that enable us substantially enhance the spatial and depth resolution of range data in real-time while preserving features, reducing random noise and eliminating artifacts like texture copying phenomenon. We have shown that the results of our approach exceed the reconstruction quality obtainable with related methods from the previous literature. Adapting the fastest acceleration strategy ever known and using the parallel processing power of a modern graphics processor, the construction of dynamic scene with a high resolution is feasible. In addition, the super resolution method is extended to one single TOF sensor case. Look into future, there are still rooms for improvement. For instance, some constraints and priors (e.g. gradient profile prior) are hoped to be incorporated into our algorithm for further improvement.

REFERENCES

[1] T. Ogger, K. Griesbach, et al., "3D-Imaging in real-time with miniaturized optical range camera," In Proc. OPTO, pp. 89-94, 2004.

[2] SwissRanger™ SR-3000, MESA Imaging inc. http://www.mesa-imaging.ch.

[3] Q.Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and Occlusion Handling," IEEE Trans. PAMI, vol. 31, no. 3, pp. 492-504, 2009.

[4] C.-K. Liang, C.-C. Cheng, Y.-C. Lai, L.-G. Chen, and H. Chen, "Hardware efficient belief propagation," In Proc. CVPR, pp. 80-87, 2009.

[5] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," In Proc. NIPS, 2005.

[6] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," In Proc. CVPR, pp.1-8 , 2007.

[7] C. Tomasi, and R. Manduchi, "Bilateral filtering for gray and color images," In Proc. ICCV, pp. 839–846, 1998.

[8] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," ACM

Transactions on Graphics (TOG)，vol. 26, no. 3, pp. 96:1–5, 2007.

[9] Nvidia Corporation. CUDA: compute unified device architecture programming guide. Technical report, 2008.

[10] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "High-quality scanning using time-of-flight depth superresolution," In Proc. CVPRW08, pp. 1-7, 2008.

[11] A. K. Riemens, O. P. Gangwal, B. Barenbrug, and R.-P. M. Berretty, "Multistep joint bilateral depth upsampling," In SPIE Vol. 7257: Proc. VCIP,2009.

[12] F. Porikli, "Constant time O(1) bilateral filtering," In Proc. CVPR, pp. 1-8, 2008.

[13] J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," ACM Transactions on Graphics (TOG), vol. 26, no. 3, pp. 103:1-10, 2007.

[14] Q. Yang, H.-H. Tan, and N. Ahuja, "Real-time O(1) bilateral filtering," In Proc. CVPR, pp. 557-564, 2009.

[15] M. Böhme, M. Haker, T. Martinetz, and E. Barth, "Shading constraint improves accuracy of time-of-flight measurements," In Proc. CVPRW08, 2008.

[16] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noise-aware filter for real-time depth upsampling," In M2SFA208: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, 2008.

[17] B. Weiss, "Fast median and bilateral filtering," In: Siggraph, vol.25, pp. 519–526, 2006.

[18] S. K. Gehrig, and U. Franke, "Improving stereo sub-pixel accuracy for long range stereo," In Proc. ICCV, pp. 1-7, 2007.

[19] M. Shimizu, and M.Okutomi, "Precise sub-pixel estimation on area-based matching," In Proc. ICCV, pp. 90–97, 2001.

[20] R. Szeliski, and D. Scharstein, "Sampling the disparity space image," IEEE Trans. PAMI, vol. 26, no.3, pp. 419–425, 2004.

[21] J. Zhu, L. Wang, R. Yang, J. Davis, "Fusion of time-of-flight depth and stereo for high accuracy depth maps," In Proc. CVPR, 2008.

[22] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," Opt. Lett., vol. 33, pp. 156-158, 2008.

Figure 4. Super resolution result on Middlebury datasets. (a) Before refinement. (b) Using Diebel's approach [5].(c) Using Yang's approach [6]. (d) Using our approach.