A fast adaptive guided filtering algorithm for light field depth interpolation

Chen, Jie; Chau, Lap-Pui

2014

Chen, J., & Chau, L. P. (2014). A fast adaptive guided filtering algorithm for light field depth interpolation. 2014 IEEE International Symposium on Circuits and Systems (ISCAS), 2281-2284.

<https://hdl.handle.net/10356/100768>

<https://doi.org/10.1109/ISCAS.2014.6865626>

© 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [<http://dx.doi.org/10.1109/ISCAS.2014.6865626>].

Downloaded on 26 Apr 2023 21:32:21 SGT

A Fast Adaptive Guided Filtering Algorithm for Light Field Depth Interpolation

Jie Chen, Lap-Pui Chau Nanyang Technological University, Singapore, 639798

Email: {jchen5, elpchau}@ntu.edu.sg

Abstract

Light field camera provides 4D information of the light rays, from which the scene depth information can be inferred. The disparity/depth maps calculated from light field data are always noisy with missing and false entries in homogeneous regions or areas where view-dependant effects are present. In this paper we proposed an adaptive guided filtering (AGF) algorithm to get an optimized output disparity/depth map. A guidance image is used to provide the image contour and texture information, the filter is able to preserve the disparity edges, smooth the regions without influence of the image texture, and reject the data entries with low confidence during coefficients regression. Experiment shows AGF is much faster in implementation as compared to other variational or hierarchical based optimization algorithms, and produces competitive visual results.

I. INTRODUCTION

The recent development of light field imaging technologies promises new possibilities to the conventional imaging industry [1]. With the extra dimensional light ray information captured by various camera structural designs (eg. inserted micro-lens array, multi-camera array, coded aperture techniques etc.), applications such as scene depth inference, after-shot refocusing, 3D scene reconstruction etc., can be implemented.

The challenge of depth inference from light field is similar to that of traditional stereo vision, however instead of a pair of input images, the light field extends the disparity space to a continuous or multiple discrete ones. Bishop et al. [2] implemented an iterative algorithm to estimate the scene depth by iteratively searching and filtering among multiple aliased views for a best correspondence match.

As the dimension at the aperture plane are usually densely sampled, this enables the formation of an epipolar-plane image (EPI) based on which the correspondence problem can be greatly

relieved, as the corresponding scene points will linearly line up as tilting straight lines in the EPI. By finding their slopes, the depth map can be calculated. Wanner et al. used a structure tensor based on local gradients to estimate the direction of lines on EPI [3]. The tensor is easy and highly efficient to implement. Kim et al. [7] proposed a scoring mechanism for all the hypothetical disparities for each scene point, and choose the highest scored disparity as their estimation. This method proves to be more accurate than the previous however slightly more time and memory consuming.

Either based on correspondence matching or EPI based slope calculation, the output depth map estimation is expected to be noisy due to several reasons such as the lack of matching clue for homogeneous regions, view-dependant effects such as occlusion or specularly. Some sort of filtering or optimization is always need for a refined output, see examples such as [4]–[6]. Notably, Wanner et al. [3] used a energy minimization approach combined with state-of-the-art convex optimization algorithms; they produced high quality depth maps, however at the cost of very high computational complexity. A “fine-to-coarse” approach was implemented in Kim’s recent work [7], where disparity estimations from lower resolution EPIs were used to interpolate the missing data in higher resolution EPIs due to lack of estimation confidence. Kim’s method produces state-of-the-art quality depth map outputs, with efficient implementation (in an order of 1 minute for a 2M pixel image).

The concept of “Guided Filtering” was first introduced by He et al. in [8], where a guidance image was used to filter another image. In appreciation of this idea, we propose an adaptive guided filtering (AGF) algorithm for disparity/depth map optimization of light field images. The centre view from light field data is used to provide the image contour and texture information, the filter is able to preserve the disparity edges, smooth the regions without influence of the image texture, and reject the data entries with low confidence during coefficients regression. Experiment shows AGF is much faster in implementation as compared to other variational or hierarchical based optimization algorithms, and produces competitive visual results.

A. The Light Field Data and Epipolar Plane Image

A typical light field camera structure is illustrated in Fig. 1(a), where a micro-lens array is inserted into a conventional camera [1], [4]. The micro-lens diverges the focused light shaft onto the camera sensor with respect to their originating point from the main lens plane. The Sensor

Fig. 1: (a) Illustration of a typical light field camera with inserted micro-lens array. (b) Demonstration of different views at the aperture plane, and (c) an example of epipolar plane image.

records each light ray’s spatial information on both the main lens and the micro-lens array plane [2].

Multiple choices are available for the organization of light field data. In this paper, we use the 4D (two-plane) set-up, where the light field data are organized as a collection of pinhole views from different aperture (main lens) positions. Illustrated in Fig. 1 (b), we relate to the 2D aperture plane as $\Pi(s, t)$, and the 2D image plane within each view as $\Omega(x, y)$. Each light ray can be expressed as $L(x, y, s, t)$.

When we fix a horizontal line in Π by setting $t = t^*$, and in Ω by $y = y^*$, the light ray will be restricted to a subspace $\{E(x, s) | E(x, s) = L(x, y^*, s, t^*)\}$, the resulting image $E(x, s)$ is called the Epipolar Plane Image (EPI). Fig. 1 (c) shows an example of EPI from the synthesis light field image “Buddha” [9]. EPI reveals the linear spatial disparity of a scene point among different viewing directions. By detecting the slope of each tilting lines on EPI, a disparity map, and further, scene depth map can be calculated [3], [7].

B. Disparity Inference from EPI

Wanner et al. used a structure tensor based on local gradients to estimate the direction of lines on EPI [3]. The tensor is easy and efficient to implement, however the output is very

noisy, especially for the homogeneous regions. Kim et al. proposed a scoring mechanism for all the admissible disparity values at each scene point [7], this method is slightly more time and memory consuming. We adopt Kim's method in our algorithm though, in consideration of its much better disparity output. Here we briefly elaborate this method.

For an image pixel x_m at the centre view of the light field, if given a hypothetical disparity value p for x_m , its corresponding line on the EPI should be:

$$\{(x, s) | x = x_m + (s_m - s)p, s = 1, \dots, N_s\}, \quad (1)$$

where N_s is the horizontal dimension of Π , and s_m correspond to the centre along dimension s (Fig. 1 (c)). The collection of all the EPI intensities along such line forms the set:

$$\mathfrak{R}(x, p) = \{E(x + (s_m - s)p, s) | s = 1, \dots, N_s\}. \quad (2)$$

If the hypothetical disparity value p is valid for x , with the assumption of a Lambertian surface, the intensity elements in $\mathfrak{R}$ should be the same, or closely approximate; and vice versa if p is invalid. A score is assigned to each admissible disparity values p to evaluate the proximity of intensities along the EPI line using Eqn. (3), and the disparity value with the highest score will be used as disparity output for pixel x . In Eqn. (3), $r_m = E(x, s_m)$ is the EPI radiance of the centre view at x , and H is a constant.

$$S(x, p) = \frac{1}{|\mathfrak{R}(x, p)|} \sum_{r \in \mathfrak{R}(x, p)} \left(1 - \left\| \frac{r - r_m}{H} \right\|^2\right) \quad (3)$$

$$P(x) = \arg \max_p S(x, p) \quad (4)$$

The output disparity map P based on this method is expected to be noisy because of the following reasons: First, homogeneous image regions also lead to homogeneous EPI regions, and the larger the area, the broader the error range is likely to be; Second, image noise or the failure of the lambertian scene surface assumption will produce misleading disparity clues on the EPI; Third, view-independent effects such as occlusions, or specularity.

A simple measurement is used to evaluate the confidence of disparity estimation, which is the 1D variance along the EPI image:

$$C(x, s) = \sum_{x' \in N(x, s)} \|E(x, s) - E(x', s)\|^2. \quad (5)$$

Here $N(x, s)$ is a 1D local neighbourhood centred at pixel (x, s) . C evaluates the radiance variation in the region, and represents the possibility of the correctness of the disparity estimation P .

II. PROPOSED ALGORITHM

A. Design Goals

Based on the noisy input disparity/depth map, we aim to design a filter that can: 1) When there are disparity changes, sharpen the edges based on the corresponding input guidance image edges; 2) Smooth the disparity area with low variation, resilient from influences of guidance image textures; 3) Remove the low confidence disparity values, and interpolate them with correct ones;

B. Definition of Filter Kernel

In this subsection we elaborate the kernel of the proposed filter. The major assumption of the design is: a small segmented patch of the ground truth disparity/depth map can be expressed as an affine transform of its corresponding image patch in the RGB space [8]:

$$q_i = a_k I_i + b_k, \forall i \in w_k, \quad (6)$$

where a_k, b_k are the affine transform coefficients for the patch centred at pixel k , both are constant in a patch window of size w_k . I_i is the input guidance image, which is the centre view of the light field, and q_i is the output disparity value. This linear model conforms to our expectation that depth changes only exist where RGB intensity changes also exist (as $\nabla q = a \nabla I$); while at the same time, for the textured surfaces with no disparity variations, RGB intensity variations should not always correspond to depth changes (in which case $a = 0$).

In order to decide the coefficients a_k and b_k for each image patch, we propose the following optimization target function:

$$\{a_k, b_k\} = \arg \min_{a_k, b_k} \sum_k \sum_{i \in w_k} (\beta_i (a_k I_i + b_k - p_i)^2 + \varepsilon a_k^2), \quad (7)$$

where $\beta_i (a_k I_i + b_k - p_i)^2$ is the fidelity term: p_i being the input noisy disparity value at pixel i , and β_i is the refined disparity confidence for pixel i according to:

$$\beta_i = \begin{cases} 0 & C_i \leq C_{th} \\ C_i & C_i > C_{th} \end{cases}. \quad (8)$$

With the definition of β_i , different weights will be assigned to input disparity values during the regression, based on their respective estimation confidence C_i . More confident values will be assigned with more significance during regression, and vice versa. What's more, a lowest confidence threshold is set at C_{th} , below which its corresponding disparity value will be assigned with zero weight, which means it has no impact on the coefficient regression, as they are considered as outliers.

The second term εa_k^2 is a regularization term which prevents a_k from getting too large: therefore prevents the output disparity from being over adapted to the guidance image texture [8], causing false disparity variations. ε is the parameter that controls the regularization effect. In our implementation, we set $\varepsilon = 10^{-7}$.

The target function (7) is quadratic, and the solution to it can be given by linear regression:

$$a_k = \frac{\sum_{i \in w_k} (\beta_i p_i I_i - \tilde{p}_k \beta_i I_i)}{\sum_{i \in w_k} \beta_i (I_i^2 - \tilde{I}_k I_i) + \varepsilon} \quad (9)$$

$$b_k = \frac{\sum_{i \in w_k} (p_i \beta_i - a_k I_i \beta_i)}{\sum_{i \in w_k} \beta_i} = \tilde{p}_k - a_k \tilde{I}_k \quad (10)$$

where $\tilde{I}_k$, $\tilde{p}_k$ are the confidence weighted average of the guidance image patch and disparity map patch centred at pixel k respectively.

One pixel will be included in multiple patches during regression, and the final output disparity value will be calculated as an average of the filtering output from all patches the pixel has been involved with:

$$q_i = \frac{1}{|wn_i|} \sum_{k: i \in w_k} (a_k I_i + b_k), \quad (11)$$

where $|wn_i|$ is the number of patches that includes pixel i . Because of the patch size varying mechanism we are about to introduce in the next subsection, each pixel's $|wn_i|$ value might be different.

C. The Patch Window Size Variation Mechanism

For a further improved filter performance, we vary the patch window size w_i for each pixel based on its confidence value.

$$w_i = (1 - \frac{C_i}{C_{\max}})(W_{\max} - W_{\min}) + W_{\min}, \quad (12)$$

where W_{max} , W_{min} are the minimum and maximum patch window sizes, which are set to 3 and 11 respectively in our implementation, and $C_{max} = \max_i C_i$.

The idea behind this mechanism is straightforward: high confident values doesn't need much filtering, a small patch window is useful for keeping its original value; however those values of lower confidence require a larger patch size and more neighbouring samples to provide disparity clues for replacement or interpolation.

An illustration of the proposed AGF filtering performance is provided in Fig. 2, where Fig. 2(a) is the centre view of the light field image ‘‘Maria’’ [9] (9×9 views, each view 926×926 pixels). Fig. 2(b) is a zoomed in region of its disparity map estimated using the ‘‘scoring’’ method, the dark pixels are the ones with estimation confidence lower than threshold Cth , and they have been removed before the disparity map is fed as input to AGF. Fig. 2(c) is the filtering result for the zoomed in region in (b). We can see the edges of the disparity map is preserved according to the original guidance image, the missing low confidence pixels are interpolated with filter output values. The final disparity map is visually satisfying.

We conclude this section by giving an intuitive understanding of the proposed adaptive guided filter:

Low-variance-smoothing: For pixels with smooth neighbourhood in the guidance image I_i (disparity confidence low), the regularization term εa_k^2 will force a_k to approach zero, and b_k will equal the weighted average of all the eligible disparity values within the patch, which smooths the region.

Edge-preservation: For pixels with high variation in its I_i neighbourhood (disparity confidence high), the regularization effect from εa_k^2 will diminish, disparity value will be matted according to the edges in the guidance image. Furthermore, because of the patch window size varying mechanism, the patch size will be small for these regions with high variations, thus a shaper edge is expected because of less averaging in Eqn. (11).

Outlier-exclusion: Because of the weighting parameter β_i in the fidelity term, questionable disparity entries will be ruled out automatically ($\beta_i = 0$), and more strict conformance will be given to the high confidence entries (β_i large) during regression.

Fig. 2: AGF processing result for the light field “Maria”. (b) is the zoomed in disparity map calculated based on EPI, low confidence values are removed. (c) is the AGF output disparity map.

III. EXPERIMENT RESULTS

We first conducted experiment on two sets of light field images from the dataset provided in [9]. Two images are “Maria” (9×9 , 926×926 pixels each view), and “MonasRoom” (9×9 views, each view 768×768 pixels). Fig. 3 shows the result. The first row are the centre view of the input light field images. Second row are the disparity maps that were fed into the adaptive guided filter, the pixels below the confidence threshold C_{th} have been taken out. The last row show the filtering result. We can see the interpolated area are smooth, and the edges are well preserved.

Comparison has been made between our algorithm with the Globally Consistent Labeling (GCL) method proposed in [3] on the light field image “Buddha” (9×9 views, with each view 768×768 pixels). Our output is visually competitive, while GCL reported a timing of around 10 minutes, the proposed filtering algorithm can finish within 1 second.

IV. CONCLUSION

The disparity map estimation from a Light field image is always noisy with missing and false data in homogeneous regions or areas where view-dependant effects are present. In this paper we proposed an adaptive guided filtering (AGF) algorithm to get a optimized output disparity map. A guidance image is used to provide the image contour and texture information, the filter is able to preserve the disparity edges, smooth the regions without influence of the image texture, and reject the data entries with low confidence during coefficients regression. Experiment shows

Fig. 3: AGF processing result for the light field (a) “Maria” and (b) “MonasRoom”. First row are the input images, second row are the confidence thresholded disparity maps, third row are the AGF processing result.

AGF is much faster in implementation as compared to other variational or hierarchical based optimization algorithms (at least 100 times faster than the GCL [3]), and produces competitive visual results.

REFERENCES

- [1] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” *Computer Science Technical Report CSTR*, 2005.
- [2] T. E. Bishop and P. Favaro, “The light field camera: Extended depth of field, aliasing, and superresolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, 2012.

Fig. 4: AGF processing result for the light field “Buddha”. (a) is input image, (b) is the disparity map calculated based on EPI. (c) is the processing result using GCL proposed in [3]. (d) is our result.

- [3] S. Wanner and B. Goldluecke, “Globally consistent depth labeling of 4d light fields,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012.
- [4] E. H. Adelson and J. Y. Wang, “Single lens stereo with a plenoptic camera,” *IEEE transactions on pattern analysis and machine intelligence*, 1992.
- [5] T. Stich, A. Tevs, and M. A. Magnor, “Global depth from epipolar volumes—a general framework for reconstructing non-lambertian surfaces,” in *3DPVT*. Citeseer, 2006.
- [6] T. E. Bishop and P. Favaro, “Full-resolution depth map estimation from an aliased plenoptic light field,” in *ACCV*. Springer, 2011.
- [7] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, “Scene reconstruction from high spatio-angular resolution light fields,” *To appear ACM Transactions on Graphics. (Proc. SIGGRAPH)*, 2013.
- [8] K. He, J. Sun, and X. Tang, “Guided image filtering,” in *ECCV 2010*. Springer, 2010.
- [9] S. Wanner, S. Meister, and B. Goldluecke, “Datasets and benchmarks for densely sampled 4d light fields,” in *Vision, Modeling & Visualization*. The Eurographics Association, 2013.