<table>
<thead>
<tr>
<th>Title</th>
<th>Application of preconditioned alternating direction method of multipliers in depth from focal stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Javidnia, Hossein; Corcoran, Peter</td>
</tr>
<tr>
<td>Publication Date</td>
<td>2018-04-06</td>
</tr>
<tr>
<td>Publisher</td>
<td>Society of Photo-optical Instrumentation Engineers (SPIE)</td>
</tr>
<tr>
<td>Link to publisher's version</td>
<td><a href="https://doi.org/10.1117/1.JEI.27.2.023019">https://doi.org/10.1117/1.JEI.27.2.023019</a></td>
</tr>
<tr>
<td>Item record</td>
<td><a href="http://hdl.handle.net/10379/14582">http://hdl.handle.net/10379/14582</a></td>
</tr>
<tr>
<td>DOI</td>
<td><a href="http://dx.doi.org/10.1117/1.JEI.27.2.023019">http://dx.doi.org/10.1117/1.JEI.27.2.023019</a></td>
</tr>
</tbody>
</table>
The Application of Preconditioned Alternating Direction Method of Multipliers in Depth from Focal Stack

Abstract

Post capture refocusing effect in smartphone cameras is achievable by using focal stacks. However, the accuracy of this effect is totally dependent to the combination of the depth layers in the stack. The accuracy of the extended depth of field effect in this application can be improved significantly by computing an accurate depth map which has been an open issue for decades. To tackle this issue, in this paper, a framework is proposed based on Preconditioned Alternating Direction Method of Multipliers (PADMM) for depth from focal stack and synthetic defocus application. In addition to its ability to provide high structural accuracy and occlusion handling, the optimization function of the proposed method can, in fact, converge faster and better than state-of-the-art methods. The evaluation has been done on 21 sets of focal stacks and the optimization function has been compared against 5 other methods. Preliminary results indicate that the proposed method has a better performance in terms of structural accuracy and optimization in comparison with the current state of the art methods.

1. Introduction

The compact design of mobile cameras does not allow users access to lens properties such as the aperture. By having the control over the aperture in a camera, one can control the camera’s depth of field (and the light flux entering the camera). This means the user can decide how much of an image remains in focus around an object. Fig.1 (a) shows schematically the relation between depth of focus (in image space) and depth of field (in object space). As illustrated in Fig.1 (b), small depth of field will make the main object in focus, while the rest of the image will be less sharp. A large depth of field will keep the entire image sharp throughout its depth; this concept is shown in Fig.1 (c). This adjustable aperture feature is available in DSLR cameras but smartphone cameras have a fixed aperture as they are designed for ease of portability, robustness and low cost.

Figure 1. Demonstration of the relation between F-number and depth of field
To overcome this limitation, post capture image refocusing can be employed by using focal stack. Focal stack is a collection of images with different focus points which correspond to different depth layers. The focal setting presenting the maximum sharpness of pixel $p$ corresponds to the depth of the pixel or the distance to the camera. The combination of these images can generate the extended depth of field similar to the range being generated by optical properties of the camera. The accuracy of this effect is highly dependent on the accuracy of the corresponding depth map.

A considerable amount of researches focused on depth from focus control for decades [1][2][3][4][5][6]. Most of these methods concentrated on depth from focus/defocus or depth recovery from focal stack on light field cameras [2][3][7][8]. Using light field cameras has an advantage of capturing simultaneous multiple views with variable focal points which provide more accurate information about the depth of the scene; however the images are captured in low resolution and in small aperture the value of SNR is significantly low [2]. The size of these cameras along with the mentioned challenges makes them inapplicable for handheld devices such as smartphones. Another disadvantage of light field cameras is the disability in handling occlusion due to the lack of lateral variation being captured in different viewpoints [7].

A framework to recover depth from focal stack is presented in [1] to handle images from smartphones. The focal stack is being aligned to make it as similar as a focal stack captured by a telecentric camera. Multilabel Markov Random Field (MRF) optimization is used to generate all in focus image from the aligned stack. This method works quite well for the Lambertian scenes however the optimization problem during the calibration process is highly non-convex and that makes this process considerably slow. The other problem of this method is the processing time of the non-linear least squares minimization to jointly optimize the initially estimated aperture size, focal depths, focal length and the depth map. The whole processing time of this method is ~20 minutes for 25 frames with 640x360 pixels resolution which make this algorithm almost inapplicable as a smartphone application. The depth maps generated by this method suffer from inaccurate depth values on objects surface, especially on reflective surfaces. In some cases the depth information along the boundaries of the foreground object is mixed with the values on the background and that might result an inaccurate synthetic defocus.

The complexity of the non-convex optimization is reformulated in [4] where depth from focus is presented as a variational problem by introducing a nonconvex data fidelity term and a convex nonsmooth regularization. The nonconvex minimization problem in [4] is aimed to be solved by a linearized alternating directions method of multipliers. This method has a superior performance in comparison to state of the art methods but the convergence of the optimization function happens very slowly and in high number of iterations. Also the depth map generated by this method suffers from inaccurate depth values on objects surface and missing edges and corners.

Some other approaches in this field have been proposed to facilitate the depth from focus applications by introducing coded focal stack photography [9] or coded aperture photography [10]. These methods require physical changes in the structure of the camera and yet the generated depth maps suffer from lack of structural quality.

In this paper we present a framework to compute and optimize the depth map from high resolution focal stack which can be used to produce an accurate synthetic defocus. The proposed method has several advantages in comparison to the state of the art methods, such as:

1- Fast and better convergence of the optimization function
2- Occlusion handling in the generated depth map  
3- High structural accuracy of the depth map  
4- High performance in texture-less scenes  
5- Accurate depth information along objects’ boundaries and surface

In the next section the proposed method is explained in detail. Evaluation and comparison the results are presented in Section III.

II. Proposed Method

At the initial step the value of the focus factor for each pixel is computed at every frame of the focal stack. The value of the focus factor for a pixel \((i, j)\) over all the frames in the stack is referred as focus function. The Modified Laplacian is used in this case to compute the focus function of \(\mathcal{F}_y\):

\[
\mathcal{F}_y = (|I \times C_x| + |I \times C_y|) \times m_r
\]

where the convolution masks are \(C_x = [-1, 2, -1]\) and \(C_y = C_x^T\). The mean filter mask is used as \(m\) by the radius \(r\).

The initial depth map is computed by modelling the focus function using the 3-point Gaussian distribution [11]. The algorithm relies on 3 focus factors \(\mathcal{F}_{y-1}\), \(\mathcal{F}_y\) and \(\mathcal{F}_{y+1}\). This will result in the following focus function:

\[
\mathcal{F} = \mathcal{F}_{\text{max}} \exp\left\{-\frac{(M - S)^2}{2\sigma_F^2}\right\}
\]

where \(M\) and \(\sigma_F\) are the mean and standard deviation of the Gaussian distribution. The estimates depth values correspond to the location of \(\mathcal{F}_{\text{max}}\). As long as there is a good correlation between the Gaussian model and the focus function, the depth values get more authentic. But this situation is not constant and it can be interrupted by variety of reasons such as noise. The presence of noise in the image domain can cause the focus function not to fit on the Gaussian model. That means the initial depth map is suffering from uncertain depth values. This condition becomes sever in case of small motions of the camera. Fig. 2 (b) shows the initial estimated depth map.

This problem is reformulated to a convex minimization problem to be solved by Preconditioned Alternating Direction Method of Multipliers (PADMM) [12][13]. To define the formulation of the convex problem we refer to regularization method proposed by Rudin, Osher and Fatemi (ROF) [14] which express the minimization problem as:

\[
P(x) = \frac{\mathcal{N}^2}{2} + \lambda \times K
\]

where \(\lambda\) is the regularization parameter and:

\[
\mathcal{N} = \int_{\mathcal{F}} (I(x) - t(x)) \ dx,
\]

where \(t(x)\) represents the Frobenius norm of \(x\)

and \(K\) defines the vectorial gradient as:

\[
K = \sup \{B: \tilde{G} \in (\mathcal{F}, \mathbb{R}^2)^2\}
\]

\[
B = \int_{\mathcal{F}} t(x) \text{div} \tilde{G}(x) dx = -\int_{\mathcal{F}} (\tilde{G}, Dt(x))
\]
\( K \), prevent the function to generate ringing artifacts along the edges but it can’t handle the discontinuities. It generally presents a loss of contrast which happens due to the use of \( \ell^2 \) fidelity. To overcome this issue, the ROF function is changed to a unique global minimizer by employing the vectorial \( \ell^1 \) norm fidelity term.

\[
P(x) = \mathcal{N} + 2^4 \times K \quad (7)
\]

Using the \( \ell^1 \) norm bring the options to solve non-convex optimization problems using convex optimization methods. The important advantage of using the convex optimization is that the global optimum is achievable with a high precision in a shorter computational time. It is also independent from the initialization.

Eq. 7 can be expressed as a constrained minimisation problem:

\[
(p, q) = \arg \min_{(p,q)} \{ R(p) + S(q) \text{ subject to } T(p, q) = l \} \quad (8)
\]

where \( R \) and \( S \) are convex proper functions, \( T \) denote a nonlinear operator and \( l \) is the specified function.

The augmented Lagrange function is used to solve Eq. 8 as:

\[
L_\gamma = R(p) + S(q) + \langle p, T(p, q) - l \rangle + \frac{\gamma \| T(p, q) - l \|^2}{2} \quad (9)
\]

Giving the residuals as \( r = T(p, q) - l \) and the dual variable as \( \rho \), we can express the ADMM problem as:

\[
p^{k+1} \in \arg \min_p \left\{ R(p) + \langle p^k, T(p, q^k) \rangle + \frac{\gamma \| T(p, q^k) - l \|^2}{2} \right\} \quad (10)
\]

\[
q^{k+1} \in \arg \min_q \left\{ S(q) + \langle p^k, T(p^{k+1}, q) \rangle + \frac{\gamma \| T(p^{k+1}, q) - l \|^2}{2} \right\} \quad (11)
\]

\[
\rho^{k+1} = \gamma (T(p^{k+1}, q^{k+1}) - l) + \rho^k \quad (12)
\]

By finding the linear approximation of \( T(p^{k+1}, q^k) \) and \( T(p^{k+1}, q^{k+1}) \) around \( p^k \) and \( q^k \) using the Taylor expansion, we can reduce the nonlinearity computation overhead of Eq. 10 and Eq. 11. So:

\[
T(p, q^k) \approx T(p^k, q^k)(1 + \theta_p(p - p^k)) \quad (13)
\]

\[
T(p^{k+1}, q) \approx T(p^{k+1}, q^k)(1 + \theta_q(q - q^k)) \quad (14)
\]

\[
W_k = \theta_p T(p^k, q^k) \quad \text{Notation}
\]

\[
T_k = \theta_q T(p^{k+1}, q^k) \quad \text{Notation}
\]

To convert ADMM to a preconditioned solver, we modify Eq. 10 and Eq. 11 by adding an additional proximity term as:

\[
\frac{\lambda \| p^{k+1} - p^k \|^2}{Z^k} \quad (15)
\]

\[
\frac{2 \lambda \| q^{k+1} - q^k \|^2}{Z^k} \quad (16)
\]

\[
\| \omega \|_Z = \sqrt{(Z, \omega)} \quad (17)
\]
where $Z$ is the positive definite matrix. So the modified Eq. 10 and Eq. 11 are:

$$p^{k+1} \in \text{arg min}_p \left\{ \frac{\lambda \| p - p^k \|_2^{2/2}}{2} + R(p) + \langle \rho^k, W_k p \rangle + \frac{\gamma \| W_k p - l + W_k p^k - T(p^k, q^k) \|_2^2}{2} \right\}$$  \hspace{1cm} (18)

$$q^{k+1} \in \text{arg min}_q \left\{ \frac{\lambda \| q - q^k \|_2^{2/2}}{2} + S(q) + \langle \rho^k, T_k q \rangle + \frac{\gamma \| T_k q - l + T_k q_k - T(p^k, q^k) \|_2^2}{2} \right\}$$  \hspace{1cm} (19)

To obtain the proximity operator, we define:

$$\zeta_1^k = \frac{Z_1^k + \gamma W_k W_k}{I} \quad (\zeta_1^k < \frac{1}{\gamma \| W_k \|_2^2})$$  \hspace{1cm} (20)

$$\zeta_2^k = \frac{Z_2^k + \gamma T_k T_k}{I} \quad (\zeta_2^k < \frac{1}{\gamma \| T_k \|_2^2})$$  \hspace{1cm} (21)

and then we can obtain:

$$p^{k+1} = \frac{(p^k - \zeta_1^k W_k \times (2 \rho^k - \rho^{k-1}))}{(I + \zeta_1^k \theta R)}$$  \hspace{1cm} (22)

$$q^{k+1} = \frac{(q^k - \zeta_2^k T_k \times (\rho^k + \gamma (T(q^k, p^k) - l)))}{(I + \zeta_2^k \theta S)}$$  \hspace{1cm} (23)

based on Eq.21 and Eq.22, the proximity operator can be defined as:

$$\frac{\omega}{(I + \alpha \theta R)} = \text{arg min}_p \left\{ 2 \alpha R(p) + \| p - \omega \|_2^2 \right\}$$  \hspace{1cm} (24)

Fig.2 (c) represents the filtered depth map by using the PADMM.
Numerical comparison of these results is a challenging task as there is no ground truth and publicly available dataset, so the depth maps are compared visually. Fig. 3 shows the generated depth maps by the proposed method, Moeller, et al. [4], Helicon Focus [25] and Zerene Stacker [26]. Fig. 3 (a) shows the case that the depth maps computed by Moeller, et al. [4], Helicon Focus [25] and Zerene Stacker [26] are missing a corner of an object and some parts of the background depth information are mixed with foreground depth values. Fig. 3 (b) illustrates the scenario where the depth maps by Moeller, et al. [4], Helicon Focus [25] and Zerene Stacker [26] are suffering from inaccurate depth values on an object’s surface. Also similar to previous example, the background depth information is mixed with foreground depth values. Fig. 3 (c) represents the case where the depth maps by Moeller, et al. [4], Helicon Focus [25] and Zerene Stacker [26] are not following the edges on the object’s boundary. This might cause a problem in segmentation and synthetic defocus application. To find more visual results and the higher resolution version of the images presented in Fig. 3 please refer to Appendix 1 in supplemental material.

To determine the performance of the generated depth maps for synthetic defocus applications, we applied hexagon shaped uniform distributed blur, based on the depth layers. Fig. 4 illustrates the synthetic defocus generated based on the depth maps presented in Fig. 3 (b). Frontal object and the background are chosen as 2 focal points for each sample. The synthetic defocus of Fig. 3 (a) and Fig. 3 (c) are presented in supplemental material.
At the second part of the experiment, the performance of the proposed PADMM is compared against 5 other optimization methods including Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [18], Classical Forward-Backward [19], Forward-Backward Splitting (FBS) [20], Accelerated FBS+Restart [21][22][24] and Adaptive Stepsize Selection FBS [23][24]. The mean and standard deviation of the residual norm for each optimization method are illustrated in Fig. 5. The maximum number of iterations and the regularization parameter are set to 300 and 0.7 for all the methods respectively. As shown in Fig. 5, the presented PADMM optimization method, results in lower convergence error in comparison to other methods.
The numerical information related to convergence of the PADMM and Moeller, et al. [4] is presented as decay of energy in a logarithmic form in Fig.6 (a) and Fig.6 (b) respectively. As it is shown in Fig.6 (a) the convergence of PADMM happens around the iteration 226 and it reaches 0.01 as the decay of energy, while the function presented by Moeller, et al. [4] around the same iteration reaches to the decay of 3.6 and it is still not converged. The better value of the decay of energy within the low number of iterations shows the superior performance of the proposed PADMM.

The third part of the comparison is done against the method proposed by Suwajanakorn, et al. [1]. The reason that we performed a separate comparison against this method is not having access to the code of the algorithm. The authors of [1] kindly provided the focal stacks and the depth results published in their paper. Fig. 7 illustrates the comparison of the depth maps computed by the proposed method and Suwajanakorn, et al. [1]. Fig.7 (a) represents the case where the depth map computed by Suwajanakorn, et al. [1] is suffering from inaccurate depth values on a reflective surface and some other objects’ surface while the depth map by the proposed method covered these issues. The depth map by Suwajanakorn, et al. [1] in Fig.7 (b) shows a similar issue to the previous example, uncertain depth values along an object’s edges and surface. Fig.7 (c) shows the similar issues of the reflective surfaces and inaccurate edges which have been solved by the proposed method. However the blue
highlighted part in Fig. 7 (c) illustrates the case where the proposed method computed a patch of uncertain depth values on the background level. It is also worth pointing out the advantage of the method by Suwajanakorn, et al. [1] in computing longer depth range than the proposed method.

**Conclusion**

In this paper a modified version of PADMM optimization method is proposed to perform on depth from focal stack and synthetic defocus application. The proposed method is applied on a sequence of images produces by a camera with hypothetical focus and aperture values to generate the depth map. The proposed technique satisfies the constraint of the state of the art method such as uncertain depth
values on objects’ surface, mixed depth values on different layers of background and foreground, missed depth information on an object’s boundaries which cause faulty edges and corners in depth map.

The method is evaluated in 2 parts. First the generated depth maps with the correspondent defocused images are compared against a recent studied method. 21 sets of focal stack images are used in this comparison and all the parameters are set equally in both methods.

The second part of the evaluation is done to determine the performance of the proposed optimization technique in comparison to 5 other algorithms.

The results of both parts of the evaluation show that the proposed framework and modified PADMM doesn’t have the best yet better performance than recent depth from focal stack and optimization methods.

High structural accuracy of the depth map generated by the proposed method gives the smartphone users the ability to refocus post-capture images accurately without the need to change the aperture size. The method has been implemented in Matlab R2016a on a device equipped with Intel i7-5600U @ 2.60GHz CPU and 16 GB RAM. The computational time of the modified PADMM optimization is ~1.5 second and the whole process from initializing the focal stack to final refined depth map takes ~53 seconds on an image with 1080×1080 pixels resolution. In our future work we plan to implement the proposed algorithm as a smartphone application. However despite the performance and accuracy of the studied method, there is still the computational time of this technique which has to be considered as the trade-off.

References