Image Reflection Suppression via Convex Optimization

Akhil Kumar Donka
EEMTECH02003

Nitish Kumar
EE22MTECH02005

Sourish Chatterjee
EE22MTECH02002

U Venkata Sai Anand M.
EE21MTECH14022

***Abstract*—The need for removing undesired reflections from the images taken through graph has been of great importance in the field of computer vision. The main uses of supressing images are aesthetic purposes and giving preprocessed image as input to machine learning and pattern recognition uses. Our proposal to solve the above problem is by a convex model which takes single image as input and removes reflections by using partial differential equations by gradient thresholding and is efficiently solved using Discrete Cosine transform. By testing on synthetic and real world examples we can conclude that our method achieves desired reflection suppression in dramatically less execution time.**

I. INTRODUCTION

Unpleasant reflections are common in images taken through glass. It would be ideal if such reflections could be eliminated. With the rise in popularity of portable digital devices such as smartphones and tablets, Many of these photographs are captured in regular life on tablets. A Image reflection suppression technology that is quick to respond and easy to use is critical in order to ensure that such photos may be analysed in seconds on mobile devices, with the best dereflected results being generated in real time based on the visual perception of the user.

Given an input reflection-contaminated image Y, traditional approaches that attempt to remove the reflection focus on separating the image into the transmission layer T (the true background) and the reflection layer R , i.e., the following assumption is made

$$Y = T + R$$

Here T, R are unknowns which makes it one equation two unknowns system which seems ill-posed. By compromising on some limitations , some assumptions were made on the image conditions to get a narrow solution space.

Instead of separating the image into two layers, suppressing the reflection in a single input image, as proposed in Arvanitopoulos et al. [2], is more practical. Furthermore, precise layer separation of a single image is impossible in general. Existing ways to separating layers more or less contain misclassified information, particularly when the reflection is sharp and powerful, which can result in dark dereflected outputs. The elimination of a considerable percentage of the energy that accumulates in the reflection layer causes this.

Most image reflection removal methods to date have placed an emphasis on performance in terms of dereflection quality. Furthermore, they can only handle photos of a limited size and are frequently computationally inefficient. megapixel smartphone photographs are increasingly popular nowadays, thanks to the rapid growth of portable device technologies. As a result, in order to manage huge photos, the efficiency of such systems must be enhanced. We propose a very efficient image reflection suppression strategy that can process huge smartphone photos in seconds while maintaining competitive dereflection quality when compared to state-of-the-art approaches. Fig. 1 is an example of our approach applying on a smartphone image.

II. NOTATIONS

Throughout the paper, we use bold letters such as **T, Y, K, f** to denote matrices. Plain letters with subscripts $\mathbf{T}_{m,n}$ denotes the element of **T** at the intersection of the m-th row and the n-th column. Elementwise multiplication between matrices is denoted by $\circ$ and convolution is denoted by $*$.

III. MODEL FORMULATION

First, According to Barrow and Tenebaum who first presented a linear model for an image **Y** that contains reflections as a sum of two other images (or layers) as follows:

$$\mathbf{Y} = \mathbf{T} + \mathbf{R} \qquad (1)$$

Where $\mathbf{Y} \in \boldsymbol{R}^{m \times n}$ is the observed image and **T, R** are the transmission and reflection layers respectively.

Our proposed model relies in the assumption that the camera focuses on the transmission layer more (i.e., the objects behind the glass) so that sharp edges appear mostly in this layer. On the other hand, the reflection layer (i.e., the reflection off the surface of the glass) is less in focus so that edges in this layer are mostly weaker than those in transmission layer. This is actually the practical case since the distance from the camera to the object in focus is different from the glass. Hence, we express our assumption using the following equation:

$$\mathbf{Y} = \omega\mathbf{T} + (1 - \omega) \circ (\kappa * \mathbf{R}) \qquad (2)$$

Where $\circ$ denotes element-wise multiplication, k is the blurring kernel and $*$ denotes the convolution operation. ω is a matrix that weighs the contribution of the transmission layer at each pixel. It is important to note here that for real images, ω is not usually constant, but depends on the lightening conditions and on the position of the camera relative to the image plane.

Let's assume that $W_{i,j} = \omega \; \forall \; i,j$. Even though this constant blending factor assumption can be incorrect in real- world scenario, it is a reasonable approximation that makes the problem tractable, considering that we have only one image at our disposal from which to suppress reflection artifacts.

Our second observation is that in most cases, humans have an uncanny ability to tell apart reflections, likely because we rely on several visual cues, including human visual system discounts the intensity modulations due to the reflection in the upper-right quadrant of the image. While harder to formalize, this observation helps us to choose prior term.

To account for these two observations, we build upon the successful image smoothing approach as discussed in [1] they smooth the image by imposing a constraint on the number of non-zero gradients on the output. Their approach globally eliminates a substantial number of gradients of small magnitudes while simultaneously retaining large magnitude edges. The optimization problem they solve has the form:

$$\boldsymbol{T}^{*} = arg \min_{\boldsymbol{T}} ||\boldsymbol{T}-\boldsymbol{Y}||_2^2 + \lambda C(\boldsymbol{T}), \qquad (3)$$

Where,

$$C(\boldsymbol{T}) = \#\{(i,j) \mid |\nabla_x T_{i,j}| + |\nabla_y T_{i,j}| \neq 0\} \qquad (4)$$

The combination of the data fidelity term with the l0 prior on the image gradients ensures that the algorithm removes edges in increasing order of magnitudes. The larger the regularization parameter λ is, the more gradients are removed.

The prior term $C(\boldsymbol{T})$ encourages smoothing of the image while maintaining the continuity of large structures. However, its combination with the data-fidelity term $||\boldsymbol{T}-\boldsymbol{Y}||_2^2$ eliminates most of the high frequency details from the image, which is desirable for smoothing, but not for reflection suppression. We want to not only preserve the continuity of large structures but also retain as much of the transmission layer details. We thus revisit the data fidelity term. To avoid losing important high frequency details from the image, we propose a Laplacian-based data fidelity term to modify the objective function of Eq. (3). The Laplacian of an image is defined as

$$\mathcal{L}(\boldsymbol{Y}) = \nabla_{xx}\boldsymbol{Y} + \nabla_{yy}\boldsymbol{Y}, \qquad (5)$$

Which is equivalent to a convolution with the 3×3 kernel $\boldsymbol{k}_{\mathcal{L}} = [0,1,0;1,-4,1;0,1,0]$. A fidelity term based on the Laplacian better enforces consistency in structures of line details. Then the proposed model takes the following form:

$$\boldsymbol{T}^{*} = arg \min_{\boldsymbol{T}} \|\mathcal{L}(\boldsymbol{T}) - \mathcal{L}(\boldsymbol{Y})\|_2^2 + \lambda \mathrm{C}(\boldsymbol{T}), \qquad (6)$$

The edge information of an image is obtained by applying the Laplacian operator $\mathcal{L}(.)$. In addition, an l_0 prior of the image gradient $\|\nabla \boldsymbol{T}\|_o$ is added to the objective function. It encourages smoothing of the image while maintaining the continuity of large structures. The Laplacian-based data fidelity term better enforces consistency in structures of fine details in the transmission layer compared to a more straight forward data fidelity term $\|\boldsymbol{T}-\boldsymbol{Y}\|_2^2$. The above model removes more gradients as the regularization parameter λ increases, which is the consequence of using the l_0 prior. Essentially, it sets a threshold on the gradients of the input image and removes the gradients whose magnitudes are larger than the given threshold. The gradient-thresholding step appears as a closed-form solution in each iteration of the algorithm as discussed in Eq. (12) of [2]. Similarly, we fuse this idea into our model formulation, but in a different way. Rather than solving the minimization problem and threshold the gradient from the solution, we adopt the idea from [3, 4] and put the gradient-thresholding step directly into the objective function. We hence propose the following model as:

$$\min_{\boldsymbol{T}} \frac{1}{2}\left\|\mathcal{L}(\boldsymbol{T}) - div\left(\delta_h(\nabla \boldsymbol{Y})\right)\right\|_2^2 + \frac{\varepsilon}{2}\|\boldsymbol{T}-\boldsymbol{Y}\|_2^2, \qquad (7)$$

Where, $\mathcal{L}(\boldsymbol{Y})$ is as expressed in Eq. (5) and,

$$\delta_h(X_{i,j}) = \begin{cases} \{X_{i,j}, & if \ \|X_{i,j}\|_2 \geq h \\ 0, & Otherwise \end{cases} \qquad (8)$$

The data fidelity term $\left\|\mathcal{L}(\boldsymbol{T}) - div\left(\delta_h(\nabla \boldsymbol{Y})\right)\right\|_2^2$ imposes gradient thresholding step on the input image **Y** before taking divergence of $\nabla\boldsymbol{Y}$**.** The gradients whose magnitudes are less than h will become zero. Since data fidelity term only contains a second order term of the variable **T,** the second term $\frac{\varepsilon}{2}\|\boldsymbol{T}-\boldsymbol{Y}\|_2^2$ is added to guarantee the uniqueness of the solution, where ε is taken to be a very small value so as not to affect the performance of the data fidelity term.

IV. SOLVING THE MODEL

The proposed model as in Eq. 7 is convex with respect to the target variable **T.** Therefore, the optimal solution can be obtained by solving a system of equations, which guarantees the optimality of the solution and contributes to the fast execution time compared to iterative methods that are common among existing approaches.

The gradient of the objective function in Eq. 7 is given by:

$$\nabla_T = \mathcal{L}\left(\mathcal{L}(\boldsymbol{T}) - div\left(\delta_h(\nabla \boldsymbol{Y})\right)\right) + \varepsilon(\boldsymbol{T}-\boldsymbol{Y}) \qquad (9)$$

Let the gradient be zero, we obtain the following equation:

$$(\mathcal{L}^2 + \varepsilon)\boldsymbol{T} = \mathcal{L}\left(div\left(\delta_h(\nabla \boldsymbol{Y})\right)\right) + \varepsilon\boldsymbol{Y} \qquad (10)$$

This equation is a variation of 2D Poisson's equation. We associate it with Neumann boundary condition since we assume a mirror extension at the boundary of the image, which implies zero gradient on the boundary. This boundary value problem can hence be solved via *Discrete Cosine Transform (DCT).* Let F_c, F_c^{-1} denote the two-dimensional DCT and its inverse. We introduce the following result.

Theorem: *The discretization of 2D Poisson's equation*

$$\mathcal{L}(\boldsymbol{T}) = f \qquad (11)$$

with Neumann boundary condition on an $M \times N$ grid is solved by

$$T_{m,m} = F_c^{-1}\left(\frac{|F_C(f)|_{m,n}}{K_{m,n}}\right) \qquad (12)$$

Where $\boldsymbol{T}, \boldsymbol{f}, \boldsymbol{K} \in R^{M \times N}$.
$K_{m,n} = 2\left(\cos\left(\frac{m\pi}{M}\right) + \cos\left(\frac{n\pi}{N}\right) - 2\right)$. $0 \leq m \leq M - 1, 0 \leq n \leq N - 1$.

The proof of the above conclusion is discussed in [5]. Essentially it says that after taking DCT, the left side of Eq. (11) becomes elementwise multiplication, i.e., $F_c(\mathcal{L}(\boldsymbol{T})) = \boldsymbol{K} \circ F_c(\boldsymbol{T})$ so the above conclusion follows. It is worth mentioning that the solution (12) has a singularity at (m, n) = (0, 0). To guarantee a unique solution, extra condition (for example, the value at $T_{0,0}$) must be specified beforehand.
We apply the above theorem to solve Eq. (10) notice that after taking DCT on both sides, the equation becomes

$$(\boldsymbol{K} \circ \mathbf{K} + \varepsilon\boldsymbol{E}) \circ F_c(\boldsymbol{P}) \qquad (13)$$

Where $\boldsymbol{P} \in R^{M \times N}$ demotes the right-hand side of Eq. (11) and $\boldsymbol{E} \in R^{M \times N}$ is a matrix of all 1's. Therefore, the solution to Eq. (11) is

$$T_{m,m} = F_c^{-1}\left(\frac{|F_C(\boldsymbol{P})|_{m,n}}{K_{m,n}^2 + \varepsilon}\right), \qquad (14)$$

where $K_{m,n}$ is the same as discussed in the above theorem. The uniqueness of the solution is automatically guaranteed because of the presence of ε in the denominator, which is the consequence of adding the $\frac{\varepsilon}{2}$ term in Eq. (7). Hence our algorithm is summarized as follows:

Input: **Y**, h, ε

Return $T_{m,m} = F_c^{-1}\left(\frac{|F_C(\boldsymbol{P})|_{m,n}}{K_{m,n}^2 + \varepsilon}\right)$

Output: **T**

V. EXPERIMENTS

Experiments are implemented using MATLAB 2020a on a PC with AMD Ryzen 5 5600H 3.30 GHz CPU and 8.00GB memory. Implementation of the code is done in MATLAB as per the direction given by Yang Yang et al. The approach is tested against some sample photos given by the author. We used 3 images from the dataset provided by authors at various values of h. The result and the original image is attached in below figures.

VI. IMAGES

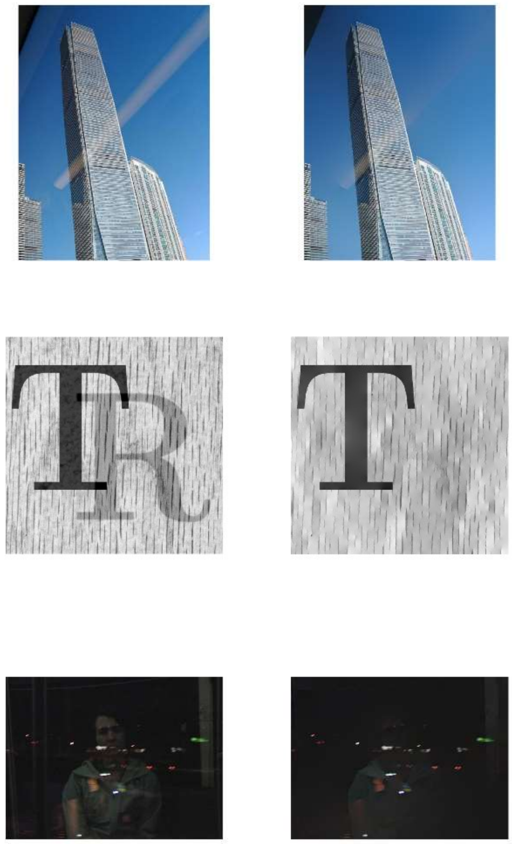

Fig. 1. Observe how a real-world image taken through the window on a train. Notice the reflection of the seat and the lights in the train. The result after the reflection suppression by our proposed method. Image size: 1080 × 1440. Execution time: 1.15s.

VII. REFERENCES

[1] L. Xu, C. Lu, Y. Xu, and J. Jia. Image Smoothing via L0 Gradient Minimization. ACM Transactions on Graphics, 30(6):174:1–174:12, 2011

[2] N. Arvanitopoulos Darginis, R. Achanta, and S. Susstrunk. ¨ Single image reflection suppression. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), number EPFL-CONF-227363, 2017.

[3] W. Ma, J. M. Morel, S. Osher, and A. Chien. An l1-based variational model for retinex theory and its applications to medical images. In CVPR, 2011.

[4] W. Ma and S. Osher. A tv bregman iterative model of retinex theory. Inverse Problem and Imaging, 6(4):697–708, 2012.

[5] W. H. Press. Numerical recipes 3rd edition: The art of scientific computing. Cambridge university press, 2007.