School of Computing Science
Simon Fraser University

Professional Master′s Program
(Visual Computing)

Co-op Technical Report

Student Name:	Suhong Kim
SFU ID:	301383275
Co-op Job Title:	Research Assistant
Company Name:	SFU
Supervisor Contact Info:	mhefeeda@sfu.ca
Co-op Start & End Dates:	May 13 – Aug 30
Semester & Year:	Summer, 2019

1. ABSTRACT

In this report, we suggest a solution to remove reflections, which occludes the contents of the video. Our algorithm generates two video sequences (background and reflection) from one input video, which shows good quality of separation results even when the video has strong reflections. In introduction, the purpose of this project and the previous works will be demonstrated. Then, we will define our model and show our approach to solve the problem. The outputs will be presented in the Results section. Finally, our future steps will be explained in the Conclusion.

2. INTRODUCTION

Taking videos with smart phones is an ordinary experience these days. However, the quality of video might not be satisfied because of reflections. For instance, if we want to capture beautiful scene out of the window during a bus trip, the taken video might have indoor reflections as well. Some professional photographers can handle those problems with the special equipment such as polarized lenses, but it is not easy to be accessed for everyday users.

Xue et al.[2] suggested the obstruction-free photography using multiple image sequence. Their optimization algorithm worked for both reflection and obstruction with minor adjustments. However, there work produced single image decomposition from multiple images. For the video reflection removal task, Nandoriya et al. [1] estimates n paired decomposition (n background frames and n reflection frames) from n frame inputs using their spatio-temporal optimization. Also, it reduces video artifacts such as temporal flickering and incomplete reflection removal.

In this project, we introduced the user intervention strategy to overcome some incomplete separation problem. Also, the stability of the optimization algorithm is improved to maintain the quality of separation for various dataset.

3. Project Description

This project aims to separate each video frame into two layers: Background layer(B) and Reflection layer(R). Based on the problem definition below, we will present our approach to generate the estimated B and R, which are showed in the result section.

3.1 Problem Definition

Figure 1 shows the image formation model, which assumes that the image captured from a camera has a linear combination of reflection object R and background scene B [2]. This model can be extended to our video sequence using

$$I_t = R_t + B_t$$

, where t is the frame number in the video. As the previous work[1] used the observation that video frames can be warped on another, we can define our model using motion field as below,

$$I_t = W^R_{t,\rho} \cdot R_\rho + W^B_{t,\rho} \cdot B_\rho$$

where $W^F_{t,\rho}$ is the motion field from frame ρ to t. However, since there are more unknowns than known in the equation, this model is severely ill-posed, which cannot be directly solved. Thus, our goal is to estimate unknowns in the above equation to recover B and R for each frame.

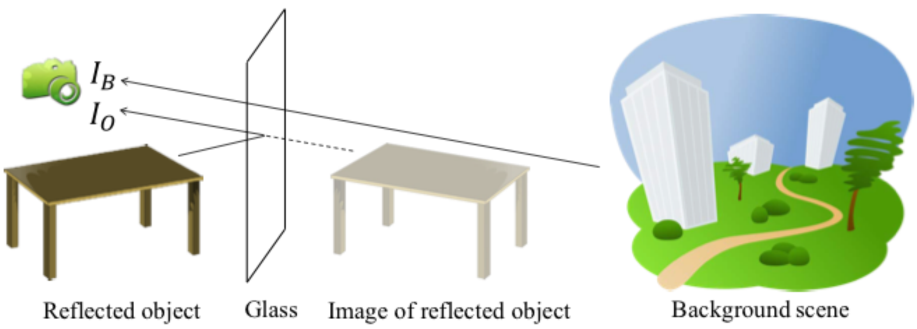

Figure 1: The image formation model from [2]. Our goal is to separate two layers (reflection object and background scene) from a single image

3.2 Our Approach

Figure 2 present our proposed method. We separate motions into two clusters based on the magnitude of velocity. Then, we estimate the initial background layer of the reference frame, which can be transformed to generate other frames' initial background layers. Reflection layers can be obtained from residual. With our optimization process, we can successfully separate two layers from each frame input.

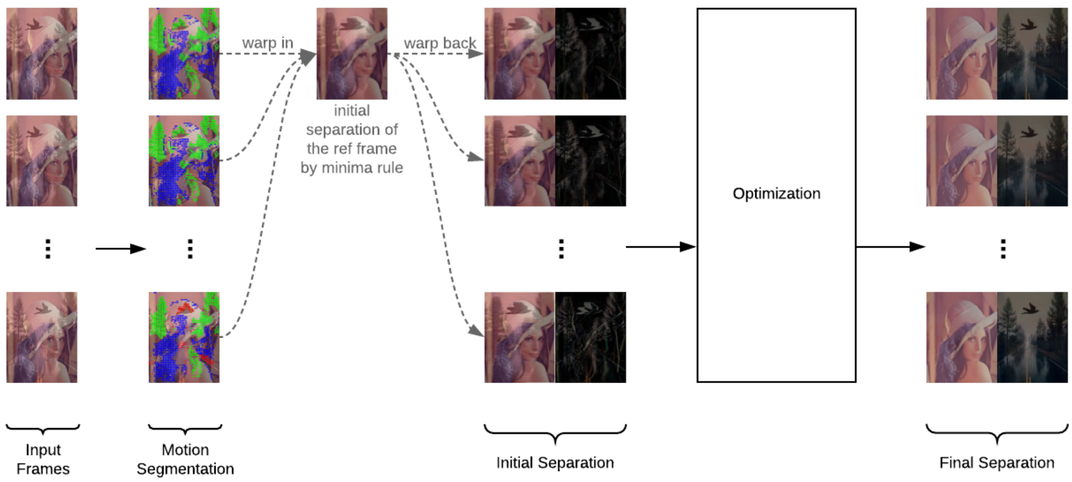

Figure 2: A Diagram of Video Reflection Removal Process

Motion Segmentation
To estimate motion fields for video frames, we track features throughout all video sequence, which are called feature point trackers. Then, they are clustered into two group using k-means method with respect to motion parallax, which assumes that background layer has more stable motion than reflection layer because reflection is much closer to the camera. Two warping matrices for each layer are estimated based on the clustered point trackers.

Initial Separation
With the estimated warping matrices in the previous step, each frame is warped into the reference frame to get the initial background layer by obtaining the minimum intensity throughout all warped frames. And then, the initial background of the reference frame can be used to generate the initial one for the other frames by warping back with the inverse transformation. Since we have initial background layer for all frames, the other reflection layers can be easily obtained by computing residual ($R_t = I_t - B_t$).

Optimization

Optimization process minimizes some of errors with respect to three error terms: data term, layer prior term and smoothness term.

$$E = \lambda_d E_d + \lambda_l E_l + \lambda_s E_s$$

$$E_d = \sum_{t=1}^{N} (||B_t - W^B_{t,\rho} \cdot B_\rho||_1 + ||R_t - W^R_{t,\rho} \cdot R_\rho||_1)$$

$$E_l = \sum_{t=1}^{N} (M_t^B \cdot |\nabla B| + M_t^R \cdot |\nabla (I_t - B_t)|)$$

$$E_s = \sum_{t=1}^{N} (|\nabla B_t| + \nabla (I_t - B_t)|)$$

For the layer prior, we applied the masks provided from the user for better separation. IRLS(Iteratively Re-weighted Least Squares) is used to solve our linear system equation with respect to minimize our error E.

3.3 Results

We tested on two types of video sequence: natural video sequence and synthesized video with ground-truth.

(a) Input **(b)** Background Layer **(c)** Reflection Layer

Figure 3: Natural Video Sequence

(a) Input **(b)** Ground-Truth(Background) **(c)** Ground-Truth(Reflection)

(d) Input **(e)** Background Layer **(f)** Reflection Layer

Figure 4: Synthesized Video Sequence

4. Conclusion

Our separation results are very successful with small artifacts. However, there are some limitations we need to improve in the next research. First, we can improve the estimation of feature point trackers to obtain better motion fields. Also, it takes a few minutes to extract one frame so that we can optimize our algorithm to reduce processing time. Finally, we can discover more ways to utilize user inputs to improve better separation.

Bibliography

[1] Ajay Nandoriya, Mohamed Elgharib, Changil Kim, Mohamed Hefeeda, and Wojciech Matusik. Video reflection removal through spatio-temporal optimization. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2411–2419, Oct 2017.

[2] Tianfan Xue, Michael Rubinstein, Ce Liu, and William T. Freeman. A computational approach for obstruction-free photography. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 34(4):79:1–79:11, Jul 2015.