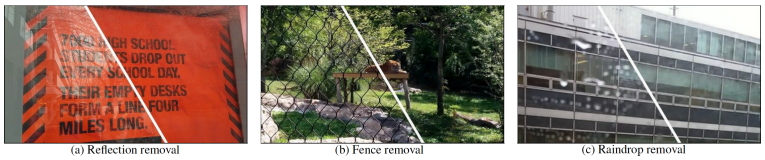
APPM X720 Biweekly Report 1

Eugene Miller

September 7th, 2020

Paper Review
Y.-L. Liu, W.-S. Lai, M.-H. Yang, Y.-Y. Chuang, and J.-B. Huang, "Learning to See Through Obstructions," arXiv:2004.01180 [cs], Apr. 2020, Accessed: Sep. 05, 2020. [Online]. Available: http://arxiv.org/abs/2004.01180.

1 Introduction



(a) Reflection removal (b) Fence removal (c) Raindrop removal

Figure 1: Seeing through (a) windows (b) fence (c) raindrop

One of the areas that has seen the most progress in applied deep learning is in image manipulation. Sometimes when we capture images of the real world, there can be artifacting, blurring, reflections or other unwanted contamination in the image. Liu et al. (2020) uses a deep convolutional neural network to remove unwanted obstructions and reflections from images.
The authors use a multi frame approach to calculate optical flows and then separate background and reflection/obstruction layers from the image sequence. Utilizing an image sequence with some movement allows the determination of motion difference between the background and obstruction. The authors use a learning based approach to reconstruct moving images into a final product, this has the benefit of not relying upon brightness or flow constancy assumptions. The final method incorporates both optimization and learning based methods but is purely data driven. This allows the method to accommodate for unforeseen violations of assumptions. It would be interesting to see how these methods could be applied as a prior to data used in computer vision research, as optical clutter and messy images can impact algorithm performance.

2 Method

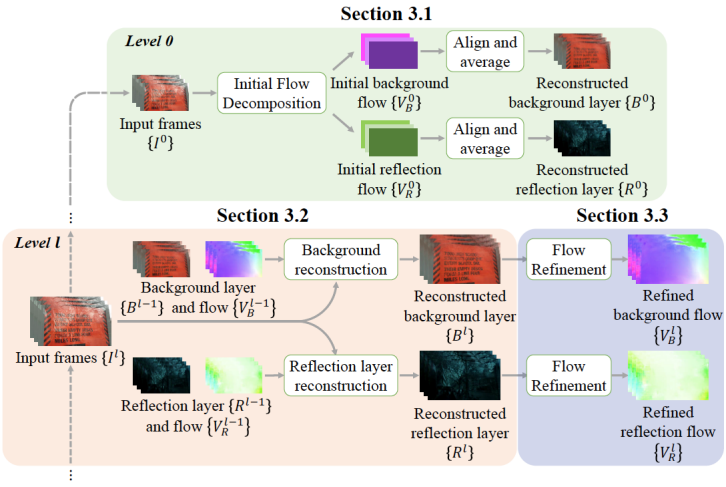


Figure 2: Algorithmic Overview

To determine the background and obstruction layers for each image of the sequence, both the motion path (flow decomposition) and layer reconstruction must occur. This is especially difficult because these are interrelated problems. Without a good layer separation, motion path is hard to determine. Without a good flow decomposition, layers are difficult to reconstruct due to misalignment. The proposed method by the authors first tackles flow decomposition, then reconstructs layers and finally refines optical flow. It is not clear whether the order of flow decomposition and layer reconstruction is arbitrary or necessary for the method.

To avoid dealing with flow fields, the authors used a uniform motion vector for each layer. This simplifies the task greatly. A flow decomposition network is formed to approach the task and consists of a feature extractor and layer flow estimator. This estimator uses global average pooling and fully connected layers to determine the motion vectors for the background and obstruction layers. The network is less complex than the layer reconstruction networks, as it is tasked with only generating vectors as opposed to images.

To generate the separate background and obstruction images, two independent networks are trained. This has the benefit of simplicity, but at the detriment of potential overlap between the output of the two networks. Ideally there should be no intersection between the features of the background and obstruction images. These networks share the same architecture but not parameters. The authors detail their layer reconstruction approach, which uses difference maps, background registered frames, invalid masks and reconstructed

background/obstruction frames. It is possible that the method could be improved using cooperatively operating networks for reconstruction of background and obstruction, but it would involve more complexity than the demonstrated method.
After reconstruction, a pretrained network (PWC) is utilized to estimate the flow fields between background reconstructions.
For training, a two-stage procedure is implemented. The first stage is as follows:
$L_{dec} = \sum_{k=1_{j\neq k}}^{T} \sum_{j=1_{j\neq k}}^{T} ||V^0_{B,j\to k} - PWC(\hat{B}_j, \hat{B}_k)\downarrow^{2^L}||_1 +$
$||V^0_{R,j\to k} - PWC(\hat{R}_j, \hat{R}_k)\downarrow^{2^L}||_1$
Where $\downarrow$ is the bilinear downsampling operator. $\hat{B}$, $\hat{R}$ are the ground truth background and reflection layers respectively. The initial flow decomposition network is then frozen and the layer reconstruction networks are trained with:
$L_{img} = \frac{1}{T\times L}\sum_{t=1}^{T}\sum_{l=0}^{L}(||\hat{B}^l_t - B^l_t||_1 + ||\hat{R}^l_t - R^l_t||_1)$
$L_{grad} = \frac{1}{T\times L}\sum_{t=1}^{T}\sum_{l=0}^{L}(||\nabla\hat{B}^l_t - \nabla B^l_t||_1 + ||\nabla\hat{R}^l_t - \nabla R^l_t||_1)$
Where ∇ is the spacial gradient descent operator. Overall, loss is calculated by:
$L = L_{img} + \lambda_{grad}L_{grad}$
The authors set λ_{grad} to 1 in all of their experiments. Both flow decomposition and layer reconstruction networks were trained using the Adam optimizer.
In addition to the above, the authors used synthetically generated image sequences to have easy access to ground truth background and reflection. This seems a necessary but unfortunate step with complex image manipulations, as synthesised sequences may not have the complexity or train the robustness necessary to tackle real data. Online optimization was also used with an unsupervised warping consistency loss due to poor performance on real-world examples. The authors note that these methods can be extended to other types of obstructions not covered directly in the paper.

3 Results

Method			Background				Reflection			
			PSNR ↑	SSIM ↑	NCC ↑	LMSE ↓	PSNR ↑	SSIM ↑	NCC ↑	LMSE ↓
Single image	CEILNet [8]	CNN-based	20.35	0.7429	0.8547	0.0277	-	-	-	-
	Zhang et al. [45]	CNN-based	19.53	0.7584	0.8526	0.0207	18.69	0.4945	0.6283	0.1108
	BDN [43]	CNN-based	17.08	0.7163	0.7669	0.0288	-	-	-	-
	ERRNet [38]	CNN-based	22.42	0.8192	0.8759	0.0177	-	-	-	-
	Jin et al. [16]	CNN-based	18.65	0.7597	0.7872	0.0218	11.44	0.3607	0.4606	0.1150
Multiple images	Li and Brown [21]	Optimization-based	17.12	0.6367	0.6673	0.0604	7.68	0.2670	0.3490	0.1214
	Guo et al. [12]	Optimization-based	14.58	0.5077	0.5802	0.0694	14.12	0.3150	0.3516	0.1774
	Alayrac et al. [1]	CNN-based	23.62	0.7867	0.9023	0.0200	21.18	0.6320	0.7535	0.1517
	Ours w/o online optim.	CNN-based	**26.57**	**0.8676**	**0.9380**	**0.0125**	**21.42**	**0.6438**	**0.7613**	**0.1008**

Table 1: Quantitative comparison of reflection removal methods on synthetic sequences. Links to papers on other methods provided in citations of paper.

The authors' method performed better than all other state of the art reflection removal methods in many different metrics. In controlled sequences,

synthetic sequences, and real sequences the new method performed better than competitors. It is clear from the visual examples that the authors' method is

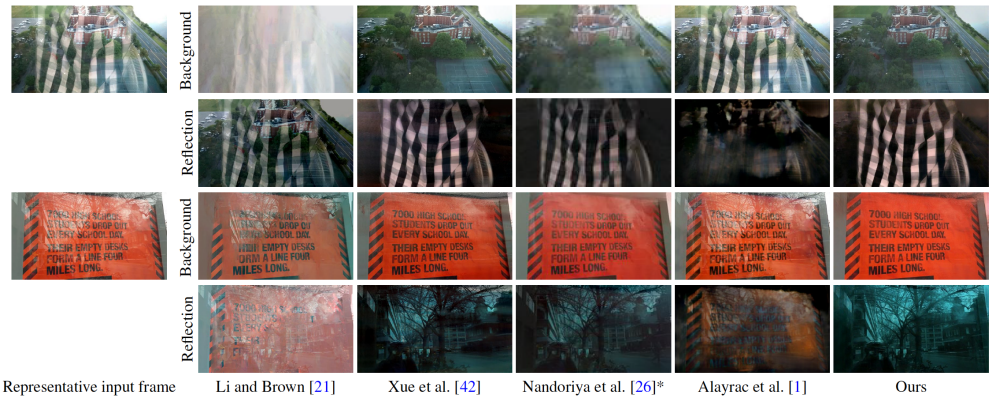


Figure 3: Visual comparison of methods

the new state of the art when it comes to obstruction and reflection removal. The authors list their key design choices as the initial flow decomposition, image reconstruction network, and online optimization. When analysing running time, they found that the method performs better than other optimization based methods when not accounting for online optimization.

The authors included also a failure case where there are two layers of reflection

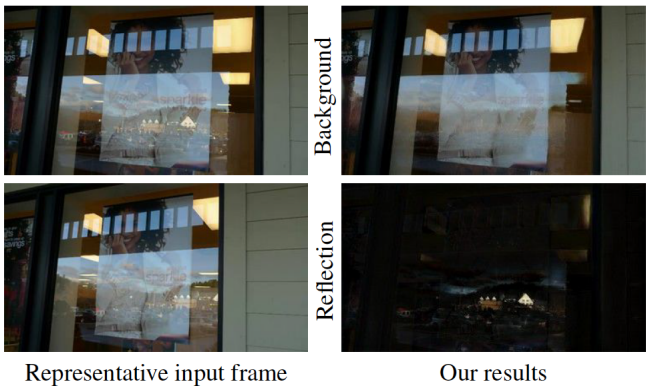


Figure 4: Failure case

present in the image. There are close reflections from the building and farther away ones from the street. The motion of the building reflection was close enough to the background that the network included it in the background reconstruction instead of the reflection. It does, however, successfully remove the

street reflections from the image. This is a possible shortcoming of the method, as it relies on the assumption that all reflections will have closer motion to each other than the background.

4 Summary

While removing obstructions and reflections from imagery remains difficult, this paper is another step along the way to full reflection and obstruction removal. Using a CNN was an important decision in the method, and it greatly improved results. The use of both learning and optimization approaches was also a major contributor in the success. Not only using learning based approaches was an intelligent way to both reduce computational overhead and apply previous optimization based research in a meaningful way. The clever use of uniform motion vectors for background and reflection layers allowed a simplification of the problem task, but it also created issues such as in the failure case. They also note that their method can be easily applied to other similar image enhancement problems, making their research all the more impressive. The methods here are accessible to students and researchers working in the field, and they will only improve from this point forward.