Stereo image de-fencing using smartphones

IIT Kharagpur

Sankaraganesh Jonna, Sukla Satapathy, and Rajiv R. Sahay

The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing

ICASSP 2017

Problem statement

Left image Right image Disparity map

Goal: To build an effecient algorithm for the detection and removal of fences/occlusions from images/videos. We solve the identified problem in three phases:

- Robust segmentation of fences/occlusions
- Optical flow between the frames
- Information fusion using TV split Bregman

Model

The degradation model for the frames of the captured video is

$$\mathbf{y}_m^{obs} = \mathbf{O}_m \mathbf{y}_m = \mathbf{O}_m \mathbf{W}_m \mathbf{x} + \mathbf{n}_m \qquad (1)$$

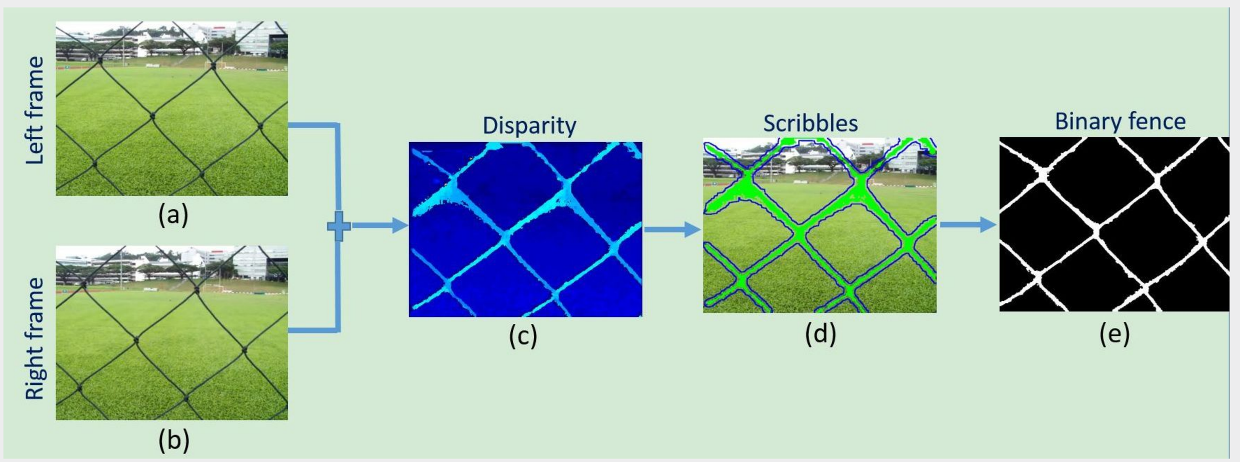
where $\mathbf{y}_m^{obs}$ is the m^{th} observation wherein the occluded pixels have been excluded using $\mathbf{O}_m$, $\mathbf{y}_m$, $m = 1, 2$ are the left and right images comprising the stereo pair, $\mathbf{O}_m$ is the fence mask corresponding to m^{th} image, $\mathbf{W}_m$ is the warp matrix, $\mathbf{x}$ is the de-fenced image and $\mathbf{n}_m$ is the Gaussian noise.

Fence detection using stereo

- Given a pair of images $\mathbf{y}_m$ (m=1,2), we want to compute the disparity map $\mathbf{D}$.
- We have exploited disparity/depth cue for the fence pattern segmentation.
- The authors in [1], trained two CNNs 'fast' and 'accurate' on pairs of small image patches.
- We employed 'fast' pre-trained model given below for matching cost computation in our work.

- Each sub-network in 'fast' CNN is made up of convolutional layers followed by ReLU layers.

Segmentation pipeline



- Exatrcted features from each one of the sub-network are used to compute cosine similarity.
- Robustly obtain the fence pixels by feeding automatically generated scribbles to mating.

Left image Right image Disparity map

Information fusion using total variation split Bregman

The de-fenced image is the solution of the following constrained optimization problem

$$\arg\min_{\mathbf{x}} \frac{1}{2} \sum_{m=1}^{p} \| \mathbf{y}_m - \mathbf{O}_m \mathbf{W}_m \mathbf{x} \|_2^2 + \mu \| \mathbf{d} \|_1 \qquad s.t.\ \mathbf{d} = \nabla \mathbf{x}$$

where p is the number of frames chosen from the video and μ is the regularization parameter. We employ the split Bregman (SB) iterative framework in [2] to solve the above problem.

$$\arg\min_{\mathbf{x}} \frac{1}{2} \sum_{m=1}^{p} \| \mathbf{y}_m - \mathbf{O}_m \mathbf{W}_m \mathbf{x} \|_2^2 + \mu \| \mathbf{d} \|_1 + \frac{\lambda}{2} \| \mathbf{d} - \nabla \mathbf{x} \|_2^2$$

where λ is the shrinkage parameter. The Bregman iterates to solve the above equation are as

$$[\mathbf{x}^{k+1}, \mathbf{d}^{k+1}] = \arg\min_{\mathbf{x},\mathbf{d}} \frac{1}{2} \sum_{m=1}^{p} \| \mathbf{y}_m - \mathbf{O}_m \mathbf{W}_m \mathbf{x}^k \|_2^2 + \mu \| \mathbf{d}^k \|_1 + \frac{\lambda}{2} \| \mathbf{d}^k - \nabla \mathbf{x}^k + \mathbf{b}^k \|_2^2$$

$$\mathbf{b}^{k+1} = \nabla \mathbf{x}^{k+1} + \mathbf{b}^k - \mathbf{d}^{k+1}$$

We can now split the above problem into two sub-problems as

Sub Problem 1:

$$[\mathbf{x}^{k+1}] = \arg\min_{\mathbf{x}} \frac{1}{2} \sum_{m=1}^{p} \| \mathbf{y}_m - \mathbf{O}_m \mathbf{W}_m \mathbf{x}^k \|_2^2 + \frac{\lambda}{2} \| \mathbf{d}^k - \nabla \mathbf{x}^k + \mathbf{b}^k \|_2^2$$

This sub-problem is solved by a steepest descent method.

Sub Problem 2:

$$[\mathbf{d}^{k+1}] = \arg\min_{\mathbf{d}} \mu \| \mathbf{d}^k \|_1 + \frac{\lambda}{2} \| \mathbf{d}^k - \nabla \mathbf{x}^{k+1} + \mathbf{b}^k \|_2^2$$

The above sub-problem can be solved by applying the shrinkage operator as follows

$$\mathbf{d}^{k+1} = shrink(\nabla \mathbf{x}^{k+1} + \mathbf{b}^k, \frac{\lambda}{\mu})$$

$$\mathbf{d}^{k+1} = \frac{\nabla \mathbf{x}^{k+1} + \mathbf{b}^k}{| \nabla \mathbf{x}^{k+1} + \mathbf{b}^k |} * max(| \nabla \mathbf{x}^{k+1} + \mathbf{b}^k | - \frac{\lambda}{\mu}, 0)$$

The update for $\mathbf{b}$ is as $\mathbf{b}^{k+1} = \nabla \mathbf{x}^{k+1} + \mathbf{b}^k - \mathbf{d}^{k+1}$. We tune the parameters μ, λ to obtain the best estimate of the de-fenced image.

References

[1] J. Zbontar and Y. LeCun *Stereo matching by training a convolutional neural network to compare image patches*, JMLR (2016).

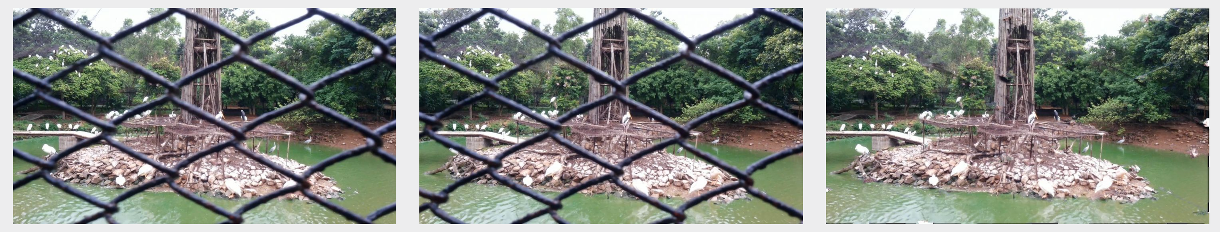
[2] T. Goldstein and S. Osher: *The split Bregman method for l1 regularized problems*, SIAM J. Imag. Sci. (2009).

[3] Y. Mu, W. Liu, and S. Yan *Video de-fencing*, TCSVT (2014).

[4] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman *A computational approach for obstruction-free photography*, TOG (2015).

[5] R. Yi, J. Wang, and P. Tan *Automatic fence segmentation in videos of dynamic scenes*, in CVPR (2016).

Experimental results



(a) (b) (c)

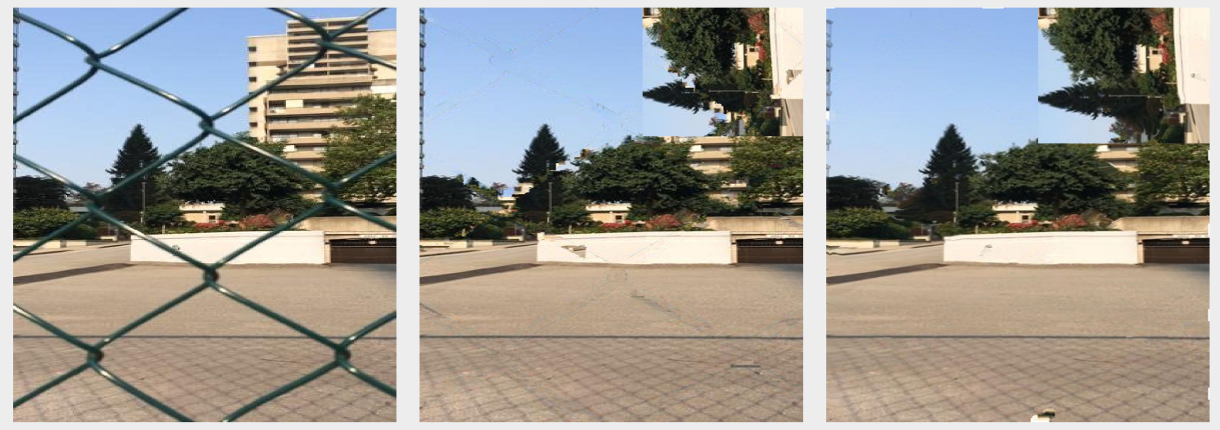
(a), (b) Two fenced observations from a video captured by us using a smartphone. (c) De-fenced image obtained by the proposed technique.

Input Mu et al. [3] Proposed

Comparison with the video de-fencing algorithm in [3].

Input Xue et al. [4] Ours

Comparison with the computational approach proposed in [4].



Input Yi et al. [5] Ours

Comparison with the state-of-the-art image de-fencing algorithm in [5].

Conclusions

- Presented a novel algorithm for fence segmentation and removal using a stereo-pair.
- We harnessed disparity cue for robust fence pixel identification.
- Computed the motion between the frames using optical flow.
- Formulated an optimization framework and the ill-posed inverse problem is solved using SB.
- We compare with several state-of-the-art works for image de-fencing.