Deep learning based fence segmentation and removal from an image using a video sequence

Sankaraganesh Jonna   Rajiv R. Sahay
Indian Institute of Technology Kharagpur, India

Krishna K. Nakka
Samsung R & D, Bangalore, India

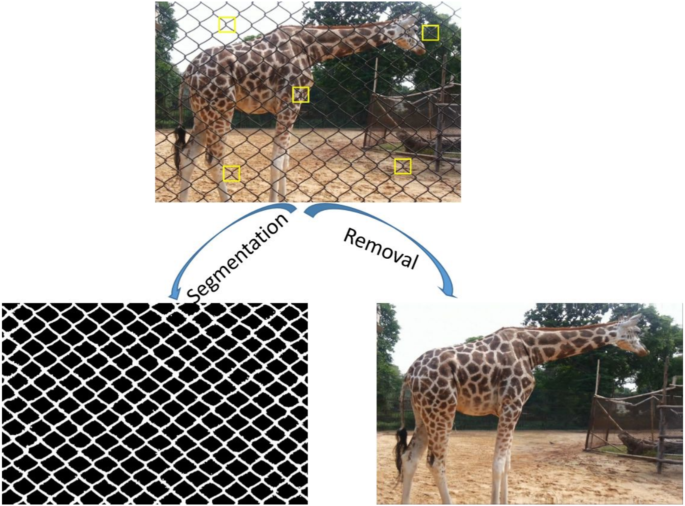
October 10, 2016

Outline

1. Introduction and Motivation
2. Prior work
3. Methodology
   - Fence segmentation
   - Optical flow
   - Optimization framework
4. Experimental results
5. Conclusion
6. References

Objectives

Segmentation and removal of fences in real-world images or videos



Introduction and Motivation

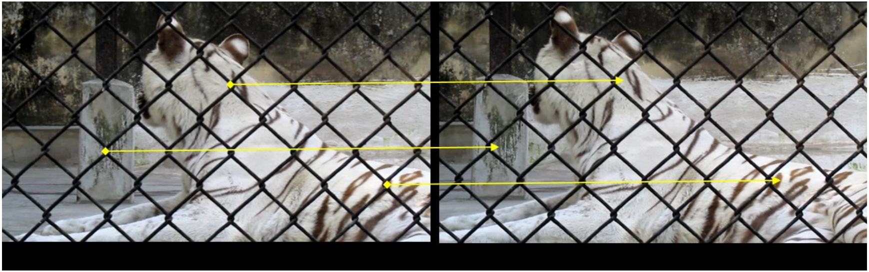
- With the advent of low-cost smartphones/phablets, the capturing of images or videos has increased manifold.
- Despite the advances in technology, sometimes the photographer is frustrated by unwanted elements in the scene.
- One such hindrance is the presence of barricades or fences occluding the object of interest.
- Security concerns have resulted in places of tourist interest being barricaded for protection.
- Therefore, a need for a post-processing tool for seamless removal of occlusions in such images exists.

Prior work

- **[Liu et al., CVPR 2008]** first addressed the de-fencing problem via inpainting of the occluded foreground pixels.
- **[Park et al., ACCV 2010]** extended image de-fencing using mutliple images, which significantly improves the performance.
- **[Mu et al., TCSVT 2014]** addressed the video de-fencing problem wherein they proposed soft fence detection method by using visual parallax as the cue.
- **[Xue et al., TOG 2015]** presented a computational approach for fence removal by panning a camera relative to the static scenes.
- **[Yi et al., CVPR 2016]** harnessed both color and motion cues for automatic fence segmentation from dynamic videos.

Cue for image de-fencing

- The basic idea is to capture a short video clip of the scene by panning the camera.
- Use a few frames from it to restore data hidden by the fence in the reference image.
- It is natural to assume that pixels occluded in the reference image are uncovered in the additional frames of the video.



Methodology

Given a video obtained by panning a camera across a static/dynamic scene occluded by a fence we propose to

1. Automatically detect spatial locations of fences or occlusions of frames in videos
2. Accurately estimate relative motion between the frames
3. Fuse data to fill-in occluded pixels in the reference image with uncovered scene data in additional frames

Degradation model

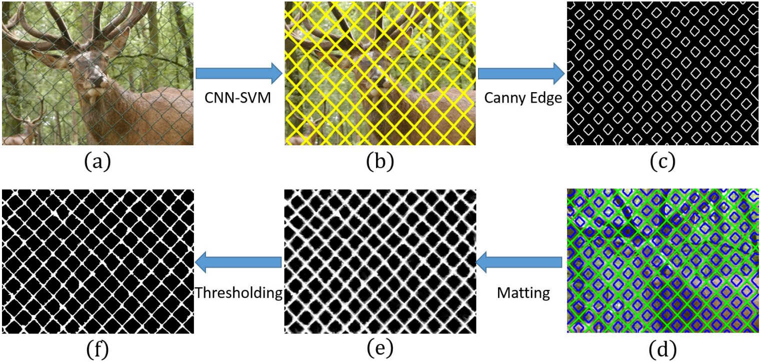
We relate the occluded image to the original de-fenced image as

$$\mathbf{O}_m \mathbf{y}_m = \mathbf{y}_m^{obs} = \mathbf{O}_m[\mathbf{F}_m \mathbf{x} + \mathbf{n}_m] \tag{1}$$

where $\mathbf{y}_m$ are observations containing fences obtained from the captured video,
$\mathbf{O}_m$ are the binary fence masks
$\mathbf{F}_m$ models the relative motion between frames
$\mathbf{x}$ is the de-fenced image
and $\mathbf{n}_m$ is Gaussian noise.

Fence Segmentation

- Automatic fence detection is the first crucial task in image de-fencing



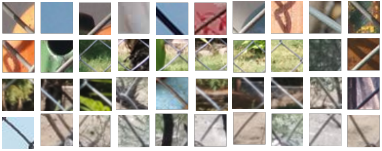
Schematic of fence mask segmentation

Pre-trained CNN-SVM

- Most of the outdoor fences are symmetry about the fence texel joints.
- To exploit this structural property of such fences, we proposed a deep learning based approach to detect joint positions
- Detected joints are subsequently connected by straight edges.
- We used pre-trained CNN as a generic feature extractor followed by a SVM classifier.

CNN-SVM (Contd.)

Texels



Non-texels

- A feature vector of dimension 4096 is extracted from an image image of size $227 \times 227 \times 3$ pixels
- SVM classifier has been trained on a dataset consists of $20,000$ fence joints and $40,000$ non-texel images
- During testing phase a sliding window is used to densly scan the test image

CNN-SVM (Contd.)

- Detected fence texels are joined by straight edges.
- Foreground scribbles are obtained by erosion operation on the connected image.
- Response obtained by Canny edge detection algorithm after dilating the preliminary fence mask and treated as background scribbles.
- We fed these automatically generated scribbles to the method of alpha matting and obtain the alpha map.
- Finally, binary fence mask is generated by thresholding alpha map.

Occlusion-aware optical flow

- Occluded image data in the reference frame is uncovered in additional frames.
- Relative motion between the frames need to estimated.
- State-of-the-art optical flow algorithms estimate the flow of visible areas between two images.
- Occlusions in images are due to depth changes, dynamic scene elements and external hindrances.
- We inpaint the motion associated with fence pixels to that of surrounding background pixel motion.

Occlusion-aware optical flow (Contd.)

- Under the incremental framework [Liu et al., PAMI 2014], one needs to estimate the best increment $d\mathbf{w} = (du, dv)$ as follows

$$\begin{aligned} E(du, dv) = \arg\min_{d\mathbf{w}} \parallel \mathbf{F}_{\mathbf{w}+d\mathbf{w}}\tilde{y}_t - \tilde{y}_r \parallel_1 \\ +\mu \parallel \nabla(u + du) \parallel_1 +\mu \parallel \nabla(v + dv) \parallel_1 \end{aligned} \tag{2}$$

- where $\mathbf{F}_{\mathbf{w}+d\mathbf{w}}$ is the warping matrix corresponding to flow $\mathbf{w} + d\mathbf{w}$, ∇ is the gradient operator and μ is the regularization parameter.

Occlusion aware optical flow (Contd.)

- Combined binary mask $\mathbf{O} = \mathbf{F}_{\mathbf{w}+d\mathbf{w}}\mathbf{O}_t || \mathbf{O}_r$ obtained by the logical OR operation between the reference and backwarped fence from the t^{th} frame.
- To estimate the optical flow increment in the presence of occlusions we disable the data fidelity term by incorporating $\mathbf{O}$ in Eq. (2) as

$$E(du, dv) = \arg\min_{d\mathbf{w}} \| \mathbf{O}(\mathbf{F}_{\mathbf{w}+d\mathbf{w}}\tilde{y}_t - \tilde{y}_r) \|_1 + \mu \| \nabla(u + du) \|_1 + \mu \| \nabla(v + dv) \|_1 \tag{3}$$

- We used conjugate gradient (CG) algorithm to solve for $d\mathbf{w}$ using iterative re-weighted least squares (IRLS) framework.

Optical flow between frames

Frame 1

Frame 2

Color coded motion between the frames 1 and 2

Optimization Framework

- We fill-in the occluded pixels in the reference image using the corresponding uncovered pixels from the additional frames.

Optimization framework (Contd.)

- We employed l_1 norm of the de-fenced image as regularization constraint in the optimization framework as follows,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left[\sum_m \| \mathbf{y}_m^{obs} - \mathbf{O}_m \mathbf{F}_m \mathbf{x} \|^2 + \lambda \| \mathbf{x} \|_1 \right] \quad (4)$$

- where λ is the regularization parameter.
- The inverse problem is solved using FISTA.

Experimental Results

- Initially, we show results obtained from our fence segmentation algorithm.
- Subsequently, we report occlusion-aware optical flow results.
- Finally, we evaluated proposed de-fencing algorithm on several video sequences and provided the comparision results.
- We ran all our experiments on a 3.4GHz Intel Core i7 processor with 16GB of RAM.

Results: fence segmentation

Input image from a video | [Park et al., PAMI 2009] | CNN-SVM

Results: fence segmentation (contd.)

Input image from a video

[Park et al., PAMI 2009]

CNN-SVM

Fence segmentation (contd.)

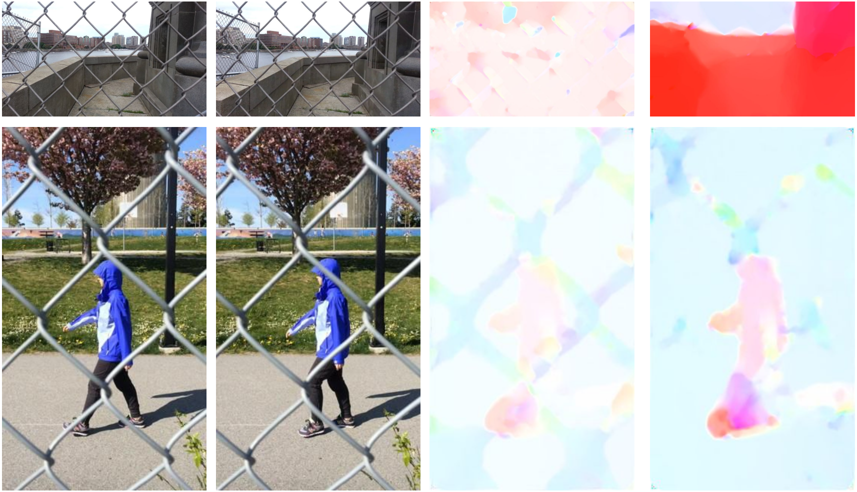
Method	NRT Database			Our Database		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Park et al., PAMI 2009	0.95	0.46	0.62	0.94	0.26	0.41
pre-trained CNN-SVM	0.96	0.98	**0.97**	0.90	0.98	**0.94**

Table: Quantitative evaluation of fence segmentation

Results: optical flow

Frame 1 | Frame 2 | [Brox et al., PAMI 2011] | Occlusion-aware optical flow

Results: optical flow (contd.)



Frame 1 | Frame 2 | [Brox et al., PAMI 2011] | Occlusion-aware optical flow

Results for image de-fencing

Input frame from a video

[Criminisi et al., TIP 2004]

Proposed de-fenced image

Results for image de-fencing (contd.)

Input frame from a video

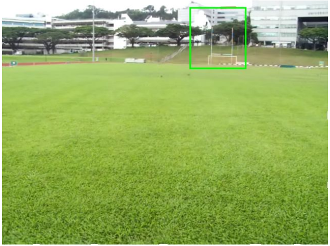
[Criminisi et al., TIP 2004]

Proposed de-fenced result

Comparison with [Mu *et al.*, TCSVT 2014]

Input video

[Mu *et al.*, TCSVT 2014]



Proposed

Comparison with [Xue *et al.*, TOG 2015]

Input video

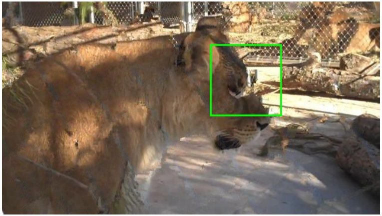
[Xue *et al.*, TOG 2015]

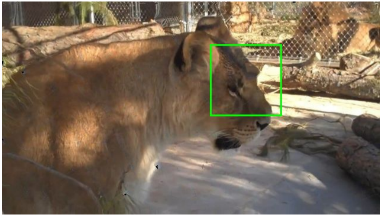
Proposed

Comparison with [Yi *et al.*, CVPR 2016]

Input video

[Yi *et al.*, CVPR 2016]

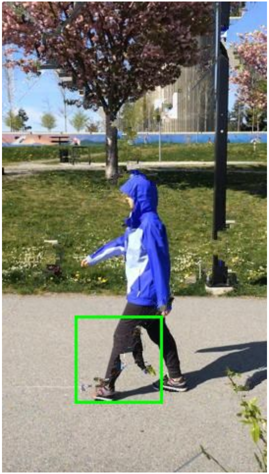




Proposed

Comparison with [Yi *et al.*, CVPR 2016]

Input video



[Yi *et al.*, CVPR 2016]

Proposed

Conclusions

- We have proposed an automated pre-trained CNN-SVM based algorithm for the first sub-problem of fence detection
- Occlusion-aware optical flow algorithm has presented under known occlusions
- An optimization-based approach was formulated for fusing data from multiple relatively shifted frames
- We compared our de-fencing results with state-of-the-art image de-fencing techniques as well as image inpainting techniques
- We believe that a real-time automatic image de-fencing algorithm will be useful especially with the advent of “smart” cameras

References I

[1] A. Barbero and S. Sra, "Fast newton-type methods for total variation regularization.," in *ICML*, pp. 313–320, Omnipress, 2011.

[2] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[3] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. ACM SIGGRAPH*, pp. 417–424, 2000.

[4] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 500–513, 2011.

[5] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. (`http://www.csie.ntu.edu.tw/~cjlin/libsvm`).

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, pp. 1097–1105, 2012.

[7] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, Nov 1998.

[8] C. Liu and D. Sun, "On bayesian adaptive video super resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 346–360, 2014.

[9] Y. Liu, T. Belkina, J. Hays, and R. Lublinerman, "Image de-fencing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1–8, 2008.

References II

[10] Y. Mu, W. Liu, and S. Yan, "Video de-fencing," *IEEE Trans. Circts. Sys. Vid. Tech.*, vol. 24, no. 7, pp. 1111–1121, 2014.

[11] M. Park, K. Brocklehurst, R. Collins, and Y. Liu, "Deformed lattice detection in real-world images using mean-shift belief propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 1804–1816, 2009.

[12] M. Park, K. Brocklehurst, R. T. Collins, and Y. Liu, "Image de-fencing revisited," in *Asian Conf. on Computer vision*, pp. 422–434, 2010.

[13] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman, "A computational approach for obstruction-free photography," *ACM Trans. Graph.*, vol. 34, no. 4, 2015.

[14] S. Jonna, K. K. Nakka, V. S. Khasare, R. R. Sahay, and M. S. Kankanhalli, "Detection and removal of fence occlusions in an image using a video of the static/dynamic scene," *J. Opt. Soc. Am. A*, vol. 33, pp. 1917–1930, Oct 2016.

[15] J. W. Renjiao Yi and P. Tan, "Automatic fence segmentation in videos of dynamic scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016.

Thank You