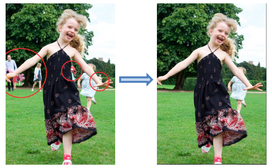
Removal of Background People Using Face Detection and Inpainting

Yuanfang Li (yli03@stanford.edu), Xin Li (xinli16@stanford.edu)
Department of Electrical Engineering, Stanford University

Motivation

- Unwanted background people in photographs is a frequent problem, particularly when taking photos at popular tourist destinations. This can be solved by identifying and removing unwanted background people and inpainting the remaining area in a realistic manner.

- Common inpainting methods require manual identification of the area to be removed. We aim to automatically identify this target area to create an end-to-end pipeline for background removal.



Related Work

- Current state-of-the-art image processing algorithms for detection and inpainting focus on deep learning methods such as convolutional neural networks (CNNs) and generated adversarial models (GANs). However these are difficult and time-consuming to train, requiring large datasets and compute power [3].

-Viola-Jones face detection algorithm[1] is a traditional image processing method able to provide fast, real-time detection of faces by using AdaBoost to combine multiple simple classifiers.

- Exemplar-based inpainting[2] uses a template window from the source region to fill in the patches of the target region in a way that propagates image structures/textures and works well for both small and large image gaps.

Reference

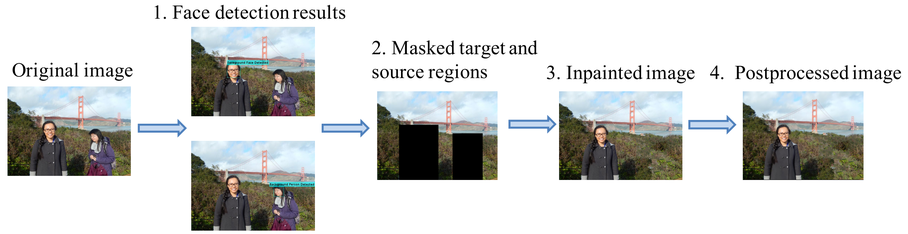
[1] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I–I. IEEE, 2001.

[2] A. Criminisi, P. Perez, and K. Toyama. Object removal by exemplar-based inpainting. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 2, pages II–II. IEEE, 2003.

[3]P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004, 2016.

Methodology

- **Pipeline**



1. Face detection using Viola Jones: Use full frontal model to detect person of interest and profile/body model to detect target person(s) to remove.

2. Mark target/source regions : Extend detected face to target region mask using standard human proportions. Mark source region as original image with target region and person of interest removed.

3. Exemplar-based inpainting:

Step 1: compute priority value

$$P(\mathbf{p}) = C(\mathbf{p})D(\mathbf{p}), \quad C(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in\Psi_{\mathbf{p}}\cap\bar{\Omega}} C(\mathbf{q})}{|\Psi_{\mathbf{p}}|}, \quad D(\mathbf{p}) = \frac{|\nabla I_{\mathbf{p}}^{\perp}\cdot \mathbf{n}_{\mathbf{p}}|}{\alpha}$$

Step 2: propagate texture

$$\Psi_{\hat{q}} = \text{argmin}_{\Psi_{\mathbf{q}}\in\Phi} d(\Psi_{\hat{\mathbf{p}}}, \Psi_{\mathbf{q}})$$

Step 3: update confidence

$$C(\mathbf{q}) = C(\hat{\mathbf{p}}) \quad \text{for } \forall \mathbf{q} \in \Psi_{\hat{\mathbf{p}}} \cap \Omega$$

4. Postprocessing with NLM filtering: Images generated by exemplar-based inpainting have patchy patterns which are hard to remove. We use NLM to remove the unwanted artifacts.

Experimental Results

- Qualitative/Quantitative results

Patch size = 17, Target region = 16.32 % of source region

Patch size = 11, Target region = 11.91 % of source region

Patch size = 7, Target region = 2.77 % of source region

Patch size = 9, Target region = 4.72 % of source region

We are able to automatically perform object removal and inpainting without manual target region selection.

- PSNR of automatic vs target selection is 26.18 dB
- Both profile and body models are required to reliably detect target regions of different sizes/postures
- Through experimentation, optimal patch size for inpainting is approximately 10% of the minimum target mask dimension

We further experiment with different source regions.

(a) (b) (c) (d)

(a) Person of interest not removed from source region
(b) Smaller patch size and local neighbourhood for inner target region
(c) Final result with detected target
(d) Result with manual target

Conclusion

We present an end-to-end pipeline for automatically detecting and removing unwanted background people from photographs that is able to achieve comparable results to hand crafted target selection. Possible future work includes generating more complex masks as well as removing patches/sudden texture changes from the inpainted area, especially for more complex backgrounds.