Removal of Background People from Crowded Scenery Image Using Target Detection and Refilling

Jiha Jang
Dept. of Electrical and Computer Engineering
Seoul National University
jeeit17@snu.ac.kr

Changhwi Park
Dept. of Geography
Seoul National University
smsychjy@snu.ac.kr

Jihyung Ko
Dept. of Computer Science and Engineering
Seoul National University
hanrista1157@snu.ac.kr

Junyul Ryu
Dept. of Computer Science and Engineering
Seoul National University
gajagajago@snu.ac.kr

Abstract

We present an application of human detection and inpainting to automatically remove unwanted background person(s) in photographs of crowded scenery images. We experiment with various object detection and targeting algorithms to optimally detect and distinguish between photograph target person and removal target(person(s) who has to be removed), produce individual masks to minimize masked background regions, as well as selecting neighborhood pixels to produce realistic refilling to cover removed pixels. The final pipeline is to be able to transform the original crowded photograph to an alone photograph of the target person and distribute the application for ordinary usages. Furthermore, we expect an application to videos.

1. Introduction

The primary motivation behind this project is to provide realistic, as well as fast application to remove unwanted background person(s) from photographs. Common approaches to this problem generally include selection of each removal target and drawing region masks manually, and choosing neighborhood regions to cover the removed region. This task becomes extremely cumbersome if the photograph contains more than a few background persons or it was taken at a very crowded location such as landmarks or tourist destinations. Thus, our aim is to (1) provide an automatic process of distinguishing 'main' photograph targets from other removal targets, (2) clean removal of the removal targets and (3) realistic repainting of the regions in order to generate desired alone photographs at the spot as the final output.

2. Methodology

2.1. Human detection and dataset generation

The very first step to take is to distinguish entire people from their background. The process begins with detecting contours of people by applying instance segmentation rather than bounding boxes that roughly encompass each person with some margin. By doing so, more surrounding features could be exploited to fill removed areas. We will suggest a supervised model that differentiates figures to be preserved from background ones to be removed. However, easily accessible datasets rarely come with a distinction between central figures and the others. By building a helper tool that generates ground truth photos under the aid of segmentation, we will compose our own dataset that can be used to train the model.

2.2. Target person distinguishment

The objective of this step is to make our model distinguish between photograph target person and removal targets. Using labeled targets from the previous step, the model is trained to narrow target selection loss in a supervised way. The work in [3] suggests a method that learns to select important image regions. Though the model in [3] has certain differences with ours in that it aims to find important image regions related to text queries, the way it calculates importance of each region can be applied to achieve our goals.

And in case there exists a difference between photograph/removal targets in frequency information due to "focusing" of the photograph, we can also apply the method from [4]. [4] suggests a method that distinguishes focused regions from defocused regions with frequency information difference obtained by applying filters to images.

2.3. Refilling algorithms for masked image region

We approach this subject in two ways: a) classical method and b) deep learning method.

2.3.1. Classical method

The work in [5] proposes a classical solution to image refilling problem by combining two classes of algorithms: (1) “textual synthesis” algorithms that generate repetitive two-dimensional textual patterns by sampling and copying color values from neighborhood region, and (2) “inpainting” that extracts linear structures(“isophetes”) and propagates it into the target region after diffusion. Although [5] focuses on removing a single - large scale object from images, we aim to extend the algorithm to remove multiple - medium/small scale objects from images by effectively partitioning the original image to regions of which contains a single object. Thus, the partitioned region will be distinguished into two sections, (1) removal target region, and (2) source region. Once these sections are determined, the region filling proceeds by recursively deciding priorities between empty pixels and updating color values to be filled. This process will iterate until all empty pixels are filled, and therefore the refilled image generated.

2.3.2. Deep learning method

To refill the erased parts, deep learning method of free-form image inpainting with gated CNN could be applied to this subject. This method aims at free-form image inpainting. Free-form image inpainting refers to a task that naturally fills multiple holes with arbitrary positions and shapes. It has two classes: (1) Gated CNN trains soft mask from data automatically. and (2) SN-PatchGAN which reduces the problem of being biased toward a specific class. For implementation, we will reference [6] implementing free-form image inpainting with gated convolution. Basically, we use a pretrained model but in case the accuracy is low, we can retrain it with labeled images.

References

[1] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE transactions on image processing*, 12(8):882–889, 2003.

[2] LI, Yaunfang and LI, Xin. Removal of Background People Using Object Detection and Inpainting. 2018.

[3] Kevin J. Shih, Saurabh Singh, and Derek Hoiem. Where to look : Focus Regions for Visual Question Answering. *CVPR*. 2016

[4] Xiaohua Qiu, Min Li, Liqiong Zhang, Xianjie Yuan. Guided filter-based multi-focus image fusion through focus region detection. Signal Processing: Image Communication. Volume 72, March 2019:35-46

[5] Criminisi, Antonio, Patrick Pérez, and Kentaro Toyama. "Region filling and object removal by exemplar-based image inpainting." *IEEE Transactions on image processing* 13.9 (2004): 1200-1212.

[6] He, Kaiming, et al. "Mask r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2017.