

# Removal of Background Crowds Using Object Detection and Inpainting

Yuanfang Li  
Stanford University  
yli03@stanford.edu

Xin Li  
Stanford University  
xinli16@stanford.edu

## Abstract

*This proposal provides an introduction of the motivation behind and final end goal of our proposed project. We aim to implement an end-to-end pipeline for identifying and removing unwanted background crowds in photographs via object detection and inpainting. We also provide a brief exploration of the existing methods that we plan to build on.*

## 1. Motivation

The primary motivation behind this project is to remove unwanted objects from photographs; we focus on removing people as unwanted background crowds is a frequent problem in photography, particularly when taking photos at popular tourist destinations. To this end, our project consists of two primary components: identifying and removing background crowds via object detection and producing a realistic final image by inpainting the remaining spaces.



Figure 1. Before/after removal of unwanted background people

## 2. Related Work

### 2.1. Object Detection

Current state-of-the-art object detection algorithms focus largely on the use of deep convolutional neural networks that require an extensive amount of data and computing resources to train but generalize well to a variety of objects. Here we focus solely on human detection, which can be done with less complex algorithms. One of the classic algorithms for object detection is the Viola-Jones face detection algorithm, which is able to provide fast, real-time detection of faces [2]. This algorithm attempts to match a set of rectangular Haar features representative of the human face to the given image. In order to perform the matching in constant time, it operates on the integral image, where each point is represented by a summation over all the pixels above and to the left of the current location. Classification is done using cascading layers employing different features, and AdaBoost is used to combine the weak classifiers provided by each feature into a single strong classifier.

Because the Viola-Jones algorithm relies on the matching of regular facial features, it can only detect faces that are directly facing the camera. In order to detect humans in a variety of poses, Viola et al propose a more complex algorithm in [3] that can be used for pedestrian detection. This algorithm uses two consecutive video frames in order to take advantage of both appearance (as in face detection) and motion information in order to detect a moving person.

### 2.2. Inpainting

Traditionally, reconstruction of holes left behind by object removal in images can be separated into two classes. Large gaps in the image can be filled in using texture synthesis algorithms, which use a sample repeating texture as the base whereas smaller gaps use inpainting, which focus more on edges and structures within the image. A method of exemplar-based inpainting is proposed in [1], which combines the two classes. Like texture synthesis, it uses a template texture window but instead of simply filling in patches from the outside inward, the patches are filled in by reverse priority value. The priority of each patch is calculated based

on how well it propagates existing edges/structures in the image as well as the confidence level in the colours assigned to the surrounding pixels.

### 3. Project Overview

#### 3.1. Goals

The final goal of this project is to remove unwanted background crowds from photos by first identifying and removing them using an object detection algorithm. We propose first identifying the person that should be kept using the classic Viola-Jones face detection algorithm. We can then use the pedestrian detection algorithm in [3] to identify and remove people moving in the background. (Note that using this method requires two captured images to produce a final image). Finally, we will use exemplar-based inpainting proposed in [1] to fill in the gaps.

#### 3.2. Milestones and Timeline

- Week 6: Gather/create dataset of consecutive images
- Week 6: Further explore the algorithms and finalize proposed image processing pipeline
- Week 7: Implement Viola-Jones face detection algorithm to identify the face/person that should be kept
- Week 8: Implement pedestrian detection algorithm to identify and remove background people
- Week 9: Implement exemplar-based inpainting to fill in image gaps
- Week 10: Prepare for the poster session and complete report

### References

- [1] 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.
- [2] 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.
- [3] P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *null*, page 734. IEEE, 2003.