Abstract

Nowadays, the enhanced capabilities of in-expensive imaging devices has led to a tremendous increase in the acquisition and sharing of multimedia content over the Internet. Despite advances in the technology of imaging sensors, many imaging conditions hamper photography in various ways. The possible common degradations are: *noise*, *poor illumination*, *rain*, *snow*, *haze*, *smoke*, *occlusions*, *missing regions*, *low-spatial resolution*, *blur*, etc. These degradations are annoying for the photographer and deteriorate the performance of many applications such as surveillance, detection, and recognition.

This thesis mainly focuses on automatic occlusion detection and removal. Occlusion segmentation is challenging due to variations in scale, shape, illumination changes, etc. Recovery of the scene from foreground occlusions is also challenging because the algorithm needs to restore faithfully missing details in the background. To achieve the above goals, we explored multiple cues for robust isolation of foreground occlusions and proposed several dis-occlusion approaches.

Initially, in the thesis it is assumed that only RGB data is available which is captured using smartphones or inexpensive cameras. To address the task of fence segmentation, we exploit the observation that real-world fences are commonly regular in shape (square or rhombic patterns) and refer to this prior assumption as *shape* cue. We also observe that foreground occlusions are almost always closer to the camera and hence, the availability of geometric information of the scene can be utilized for obstruction segmentation. The advantage of *depth* cue is that it is robust to variations in shape, illumination, or scale of the fence occlusions.

Having detected occlusions, the task of filling-in hidden areas can achieved by utilizing temporal information. Missing regions in the reference image can reappear in other frames of the video sequence due to motion parallax. Exploiting this idea, we solve the inverse problem of reconstructing the de-fenced image using smoothness prior minimizing an appropriately formulated energy function. Next, the thesis investigated the task of super-resolution image de-fencing. In order to preserve fine details in the high resolution de-fenced image, we exploit the nonlocal self-similarity property of image patches to extend the discontinuity adaptive Markov random field prior.

Since estimation of relative motion between frames of the video is important to accurately recover the scenes from occlusions, we propose to estimate occlusion-aware optical flow. Although, there have been many algorithms in the literature for depth map inpainting, most of them use the corresponding occlusion-free color

image for guidance. Since we use RGB-D data as input which is degraded by occlusions, we propose to first fill-in missing regions in the color image. Subsequently, the inpainted RGB image is used for recovering missing depth data.

In the second part of the thesis, we exploit the *blur* cue for occlusion segmentation wherein only a single degraded image is taken as input. In partially blurred images, foreground occlusions can appear out-of-focus due to the finite depth-of-field of camera. This phenomenon can be exploited to segment the occlusions irrespetive of their shape. We formulate blur detection and estimation as a superivsed learning task.

Furthermore, to address the challenge of gathering ground truth annotations for blur detection, we develop a semi-supervised learning framework using only limited amount of labelled data. Although, we segment the foreground occlusions from a single image through blur detection, filling-in of missing information is more challenging as compared to our previous multi-frame approaches. Therefore, we propose a fully-automatic generative adversarial learning algorithm to isolate the occlusions of any arbitrary shape and estimate the occlusion-free image given a single degraded photograph.

Keywords: *Inpainting, maximum a posteriori, Markov random field, super-resolution, non-local means, rgb-d data, depth completion, optical flow, total variation, low-rank regularization, space-variant blur, convolutional neural networks, generative adversarial networks, semi-supervised learning.*