Extreme View Synthesis

Kobee Raveendran

February 3, 2020

1 Summary

This paper tackles the problem of creating novel views from a sample of other viewpoints. The synthesized viewpoints could serve as good interpolations for areas in a 3D environment that the given samples cannot account for. This has been done before with good success, but this paper seeks to tackle the problem given less samples to build off of, in this case as little as two. Previous methods either had difficulty accounting for differences in depth, or if they did, produced noticeable, large artifacts (which can look like shearing or tearing in the image, unnatural distortions, or other irregularities), especially when viewpoints are extended far from the original.

2 Good points

The upsides of this paper and its method is that it has very promising visual results; compared to previous methods, this one produces results that seem sharper and approach the true viewpoint in realism. Also, the use of depth probability allows it to handle varying depth scenarios well, and perform decently even under high magnification. And finally, the authors' method fills in areas prone to occlusion very well, given that these areas are included in the views of one of the cameras (which the authors suggest the use of generative methods, such as GANs, to solve).

3 Weak points

This paper seemed very strong to me, though possibly due to the nature of the problem, I found it unsettling to see just a single quantitative analysis of their model's performance (which were average PSNR score and average SSIM). For example, even in papers doing purely generative work, such as in most GAN-based papers for computer vision, there is a good balance between qualitative benchmarks (i.e. Mechanic Turk realism studies) and quantitative benchmarks (FCN scores, Inception scores, etc.). Though I am not completely familiar with the domain of viewpoint synthesis, surely there are other quantitative metrics the authors could have used to supplement their current results and show the strength of their method.

4 Questions

When discussing Figure 6 and in their method, the authors state that they use depth probability in addition to patches from the image to fix the artifacts the occur as a result of simple image warping for viewpoint synthesis. How exactly are these patches defined and selected? How are they chosen to be varied for artifacts in different locations or of different sizes?

5 Ideas

In my current research, we are testing the use of Stereo-RCNN for 3D object detection in the autonomous driving space, which estimates and draws bounding "prisms" around objects of interest, such as motor vehicles and bicycles, given only 2D images. One problem the method attempts to solve is that of depth, which is non-trivial to extract from a 2-dimensional image. Many monocular methods attempted to use geometry to determine the depth and heading (orientation and direction) of vehicles. Stereo-RCNN and other non-monocular methods use 2 or more cameras (and thus viewpoints) to determine depth. Because of this, I believe the inputs to both Stereo-RCNN and the current paper have much in common, so in theory the method discussed in this paper could be extended or transformed to fit Stereo-RCNN, resulting in something more robust to different viewing angles and occlusion, since Stereo-RCNN can sometimes experience poor results if part of the object is occluded or at an ambiguous viewing angle.