In terms of evaluation, this paper first tests the synthesis by adding different levels of image distortion to two public datasets, and then also collects real fisheye images of three common human activities: posing, talking and walking, in a top-down viewpoint. The results obtained from the final evaluation show that the method mentioned in this paper outperforms existing methods on both synthetic and real datasets.