The impact of intermediate video frames reconstruction step on the result of 3D reconstruction of objects

Yuliya Ganeeva 1
jganeeva99@gmail.com

Vladislav Myasnikov 1
vmyas@geosamara.ru

Introduction

The task of 3D reconstruction of objects is one of the most challenging tasks in computer vision. This is usually caused by the requirement of a lot of input information, such as a large number of images or camera settings. Often the available data for object reconstruction is in the form of a limited number of images from the Internet or frames of a video sequence with a low rate. This paper estimates the performance of such widely used 3D reconstruction methods, as Akool, COLMAP, and NeRF depending on the input data. The obtained results show that the stage of intermediate frame reconstruction allows to improve the quality of 3D reconstruction in the case of a low number of input images.

The procedure for conducting experimental studies is as follows:

- Divide a 15-frame-per-second video into frames.
- Double the frame rate of video sequences by reconstructing intermediate frames of video sequences using one of the following methods: XVFI, CDFI, RRIN.
- Using the 3D reconstruction methods, reconstruct the object based on the frames obtained in step 2.
- Evaluate the result of 3D reconstruction.

3D IoU (Intersection-over-Union) is a metric used to measure the quality of the 3D reconstruction. It is calculated as follows:

$$\text{IoU}_{3D} = \frac{V_\cap}{V_\cup}$$

where $V_\cap$ is the intersection volume of the meshes, $V_\cup$ is the union volume of the meshes.

Conclusion

As a result of the study, the hypothesis was confirmed that the synthesis of frames from missing viewpoints will improve the quality of 3D reconstruction for video sequences with a low frequency describing objects. Experimentally, it was found that the frames reconstructed by the CDFI and RRIN methods are not suitable for use in the 3D reconstruction task. This is since the resulting frames have a large number of artifacts that do not allow the next stage of 3D reconstruction to correctly match frames and calculate descriptors when using the COLMAP tool, as well as when initializing the NeRF method, which is performed using the COLMAP tool. 3D reconstruction using the Akool tool using frames reconstructed by the XVFI method showed results close to those obtained on the original frames (30 FPS). However, it is worth noting that the use of the Akool tool is controversial, as the details and implementation details of the 3D reconstruction algorithm are known, which have been slightly modified and have led to the successful solution of some problems.

In the future, it is planned to develop a method for 3D reconstruction of objects based on the ideas presented in NeRF, as well as intermediate video frames reconstruction method that is resistant to camera movements that describe the shape and appearance of objects. In particular, the goal is to move away from static objects and instead focus on dynamic objects, which move.