

# Влияние реконструкции промежуточных кадров видеопоследовательности на результат 3D-реконструкции объектов

Ю.Х. Ганеева

Самарский национальный исследовательский университет  
им. академика С.П. Королева

Самара, Россия

jganeeva@gmail.com

В.В. Мясников

Самарский национальный исследовательский университет  
им. академика С.П. Королева

Самара, Россия

vmyas@geosamara.ru

**Аннотация**—Задача 3D-реконструкции объектов является одной из сложных задач в области компьютерного зрения. Как правило это связано с тем, что требуется большое количество входных данных, например изображений и параметров камеры. Часто для восстановления объектов данные имеются в виде ограниченного количества изображений из сети Интернет или кадров видеопоследовательности с низкой частотой. В данной работе производится оценка качества работы методов/инструментов 3D-реконструкции Akool, COLMAP и NeRF. Полученные результаты показали, что при недостаточном количестве входных изображений объекта, этап реконструкции промежуточных кадров позволяет повысить качество 3D-реконструкции.

**Ключевые слова**— 3D-реконструкция, NeRF, Akool, COLMAP, Alpha shapes, реконструкция кадров.

## 1. ВВЕДЕНИЕ

Задача 3D-реконструкции объектов/сцен [1] является одной из сложных задач в области компьютерного зрения. Результат реконструкции может быть использован в AR/VR приложениях, что является актуальным для создания реалистичного виртуального мира (например, Meta [2]). Для приближенной к реальности реконструкции необходимо достаточное количество изображений, позволяющее полностью описать внешний вид и форму объекта, что не всегда представляется возможным. В случае, когда требуется произвести 3D-реконструкцию динамического объекта/сцены, вопрос о количестве данных (как правило, частоте кадров видеопоследовательности или синхронизированных видеопоследовательностей) является основополагающим.

Реконструкция кадров видеопоследовательности — это процесс синтеза новых кадров видеопоследовательности на основе информации о соседних кадрах. В ситуации, когда набор данных расширяется за счёт реальных изображений с разных ракурсов – повышение качества реконструкции очевидно. В рамках исследования была выдвинута гипотеза, что синтез кадров с отсутствующих точек обзора, позволит повысить качество реконструкции. Однако, существующие лучшие среди известных (англ. state of the art, SOTA) методы реконструкции кадров по большей части не являются устойчивыми к динамике в кадре и движению камеры. В рамках работы для реконструкции промежуточных кадров видеопоследовательностей были выбраны методы XVFI [3], CDFI [4] и RRIN [5], показавшие SOTA качество реконструкции на популярном наборе данных Vimeo90K [6].

Целью данного исследования является оценка влияния этапа реконструкции кадров видеопоследовательности, описывающей внешний вид и форму объекта, на качество 3D-реконструкции. Для реализации 3D-реконструкции объектов было выбрано и оценено качество работы трёх методов/инструментов: Akool [7], COLMAP [8], NeRF [9].

Структура работы следующая. В разделе 2 приведено описание используемых для 3D-реконструкции методов. В разделе 3 представлены порядок проведения экспериментальных исследований и соответствующие результаты. Работа заканчивается заключением, в котором делаются выводы на основе результатов, полученных в ходе экспериментальных исследований.

## 2. ОПИСАНИЕ ИСПОЛЬЗУЕМЫХ МЕТОДОВ/ИНСТРУМЕНТОВ 3D-РЕКОНСТРУКЦИИ

Akool [7] – это инструмент, предназначенный для создания 3D-моделей высокого качества на основе SOTA алгоритмов искусственного интеллекта и алгоритмов SLAM [10]. Для получения 3D-модели необходимо лишь наличие RGB-изображений объекта. Все последующие вычисления производятся на удаленном(ых) сервере(ах). Результатом работы является полноценная 3D-модель (меш), описывающая форму и текстуру объекта. Стоит отметить, что предоставляется возможность одновременного создания нескольких 3D-моделей. Однако, конечная внутренняя реализация алгоритма является неизвестной.

COLMAP [8] – это инструмент, имеющий открытый исходный код и предназначенный для создания 3D-моделей из набора изображений. Результатом работы алгоритма реконструкции, реализованного в COLMAP, является облако точек. Однако полученное облако точек нуждается в постобработке, которая заключается в удалении фоновых точек с помощью инструмента MeshLab [11], и в создании на основе полученного облака точек меша с помощью алгоритма Alpha shapes [12].

NeRF [9] – это SOTA метод 3D-реконструкции объектов, который основан на синтезе новых видов для сложных сцен. Авторы получают представление сцены с помощью полносвязной (не свёрточной) глубокой нейронной сети, которая на вход принимает 5D-вектор координат (пространственные координаты  $(x, y, z)$  и радиальные координаты  $(\theta, \varphi)$ ). Нейронная сеть в результате выдает 4 числа:  $(R, G, B, \sigma)$ , где  $(R, G, B)$  – это значения для каждого цветового канала в заданной точке

пространства  $(x, y, z)$ ,  $\sigma$  – значение в диапазоне  $(0, 1)$ , показывающее есть ли в заданной точке пространства объект.

### 3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Порядок проведения экспериментальных исследований следующий:

- Разбить исходную видеопоследовательность на кадры.
- Выполнить прореживание видеопоследовательности. При прореживании из исходной последовательности кадров удаляется каждый второй кадр.
- Выполнить увеличение частоты видеопоследовательностей в два раза путем синтеза промежуточных кадров одним из методов: XVFI, CDFI, RRIN.
- Произвести реконструкцию объекта по кадрам, полученным на шаге 2.
- Оценить результат 3D-реконструкции.

Для реализации процедуры оценки реконструкции было выполнено 3D-моделирование коробки на основе предварительных замеров, а также 3D-моделирование и 3D-печать конуса. Полученные 3D-модели были приняты за эталонные и использовались далее для оценки результата 3D-реконструкции.

Оценка качества 3D-реконструкции производилась с использованием метрики 3D IoU (Intersection-over-Union), которая вычисляется следующим образом:

$$IoU_{3D} = V_{\cap} / V_{\cup}$$

где  $V_{\cap}$  - объем пересечения мешей,  $V_{\cup}$  - объем объединения мешей.

Для проведения экспериментальных исследований были записаны две видеопоследовательности с частотой кадров 30 FPS, описывающие внешний вид и форму коробки и конуса, представленных на рис. 1. При съемке объекты находились в статическом положении, камера же двигалась вокруг объектов.



Рис. 1. Объекты для реализации 3D-реконструкции: коробка, конус

Для реализации реконструкции кадров и оценки качества 3D-реконструкции из видеопоследовательностей был вырезан каждый второй кадр. В результате были получены две видеопоследовательности с частотой кадров 15 FPS.

В таблице 1 представлены лучшие результаты 3D-реконструкции с синтезированными и оригинальными кадрами.

Таблица IV. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Метод 3D-реконструкции	Метод реконструкции кадров					
	XVFI (30 FPS)		Оригинал (30 FPS)		Оригинал (15 FPS)	
	Коробка	Конус	Коробка	Конус	Коробка	Конус
Akool	<b>0.8026</b>	<b>0.7435</b>	<b>0.8548</b>	<b>0.8337</b>	0.0000	0.0000
COLMAP + Alpha shape	0.5309	0.6332	0.6700	0.8409	0.4968	0.4801
NeRF	0.1145	0.5489	0.5964	0.5516	0.0000	0.0000

### 4. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования было выявлено, что кадры, синтезированные методами CDFI и RRIN не пригодны для использования в задаче 3D-реконструкции, так как результирующие кадры имеют большое количество артефактов. 3D-реконструкция инструментом Akool с использованием кадров, синтезированных методом XVFI, показала результаты близкие к результатам, полученным на оригинальных кадрах. Полученные результаты в дальнейшем планируется использовать в задаче 3D-реконструкции динамических объектов/сцены.

### ЛИТЕРАТУРА

- [1] Гошин, Е.В. Реконструкция 3D-сцен по разноразмерным изображениям при неизвестных внешних параметрах съемки / В.А. Фурсов, Е.В. Гошин // Компьютерная оптика. – 2015. – Т. 39, № 5. – P. 770-776. DOI: 10.18287/0134-2452-2015-39-5-770-775.
- [2] Welcome to Meta | Meta [Электронный ресурс] — Режим доступа: <https://about.facebook.com/meta/> (06.02.2022).
- [3] Sim, H. XVFI: eXtreme Video Frame Interpolation / H. Sim, J. Oh, M. Kim // arXiv: 2103.16206, 2021.
- [4] Ding, T. CDFI: Compression-Driven Network Design for Frame Interpolation / T. Ding, L. Liang, Z. Zhu, I. Zharkov // arXiv: 2103.10559, 2021.
- [5] Li, H. Video Frame Interpolation Via Residue Refinement / H. Li, Y. Yuan, Q. Wang // ICASSP – IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2020. – P. 2613-2617. DOI: 10.1109/ICASSP40776.2020.9053987.
- [6] Xue, T. Video Enhancement with Task-Oriented Flow / T. Xue, B. Chen, J. Wu, D. Wei, W.T. Freeman // International Journal of Computer Vision. – 2019. – Vol. 127(8). — P. 1106-1125. DOI: 10.1007/s11263-018-01144-2.
- [7] Akool [Электронный ресурс]. — Режим доступа: <https://akool.com/> (31.01.2022).
- [8] COLMAP — COLMAP 3.7 documentation [Electronic resource]. — Режим доступа: <https://colmap.github.io/> (06.02.2022).
- [9] Mildenhall, B. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis / B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng // arXiv: 2003.08934, 2020.
- [10] Stachniss, C. Robotic Mapping and Exploration / Stachniss, C. – 2009. – Vol. 55.
- [11] Cignoni, P. MeshLab: an Open-Source Mesh Processing Tool / P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia // Eurographics Italian Chapter Conference. – 2008. – P. 129-136.
- [12] Fischer, K. Introduction to Alpha Shapes / K. Fischer. – 2009. – P. 17.