

## 3D-РЕКОНСТРУКЦИЯ ЖЕСТКИХ КОНСТРУКЦИЙ ИЗ МОНОКУЛЯРНОГО ВИДЕО

### 3D-RECONSTRUCTION OF RIGID CONSTRUCTIONS FROM THE MONOCULAR VIDEO

**I. Manshin  
G. Falkov  
S. Zuev**

*Summary.* To date, approaches based on template representations of a particular class have achieved some success. The disadvantage of such methods is the lack of flexibility of use in relation to unknown categories of objects. The paper presents a template-free approach for studying 3D shapes in one video. It uses an analysis-by-synthesis strategy that allows you to visualize the silhouette of an object for comparison with video surveillance. Without relying on a category-specific form template, the method reconstructs rigid 3D structures from videos of unknown classes.

*Keywords:* 3D shape restoration, monocular video, mash, rest shape, rendering, 3D shape.

**Маньшин Илья Михайлович**

Аспирант, Белгородский государственный  
технологический университет им. В.Г. Шухова,  
г. Белгород  
manhin@yandex.ru

**Фальков Георгий Александрович**

Аспирант, Белгородский государственный  
технологический университет им. В.Г. Шухова,  
г. Белгород  
bag.falkova@gmail.com

**Зуев Сергей Валентинович**

К.ф.-м.н., доцент, Белгородский государственный  
технологический университет им. В.Г. Шухова,  
г. Белгород  
sergey.zuev@bk.ru

*Аннотация.* На сегодняшний день подходы, основанные на шаблонных представлениях конкретного класса, достигли определенного успеха. Недостаток таких методов заключается в отсутствии гибкости использования по отношению к неизвестным категориям объектов. В работе представлен подход без шаблонов для изучения 3D-форм по одному видео. Он использует стратегию анализа путем синтеза, которая позволяет визуализировать силуэт объекта для сравнения с видеонаблюдениями. Не полагаясь на шаблон формы, специфичный для конкретной категории, метод реконструирует жесткие 3D-структуры из видео неизвестных классов.

*Ключевые слова:* восстановление 3D-формы, монокулярное видео, mash, rest shape, rendering, 3D shape.

### Введение

**В**осприятие и моделирование геометрии и динамики 3D-объектов является открытой исследовательской проблемой в области компьютерного зрения и имеет множество применений. Для восстановления 3D объекта, используя «грубый подход», необходимо подобрать такую комбинацию форм, текстур, источников освещения, чтобы она соответствовала при прямом проецировании в 2D исходному кадру данных. Такой подход обладает явным недостатком в поиске такой комбинации. Поэтому основной проблемой стоит нехватка ограничений для такой задачи [1].

Если обратиться к существующим и неплохо показавшим себя решениям, то можно заметить использование априорных данных. Такие подходы не универсальны и зачастую дорого обходятся. Например, модели направленные только на определение позы человека,

форм птиц и предметов интерьера требуют колоссальных наборов предварительно подготовленных данных [2, 3]. Такие способы не являются гибкими к появлению новых классов для 3D реконструкции.

### Описание подхода

Существует ряд исследований, которые рассматривает формы определенных категорий и изучает силуэт и ключевые точки в большой коллекции изображений для построения 3D форм из них [2, 4–8]. Однако 3D-данные, как правило, трудно получить в больших масштабах из-за конструкции датчика, хотя и возможно, благодаря вводу некоторых ограничений для конкретного рода объектов и точек взгляда на эти объекты [9, 10]. Видео служит в качестве альтернативы сканированию глубины и коллекциям изображений — видео легче получать и обеспечивает четко определенные ограничения для нескольких видов 3D-формы одного и того же экземпляра.

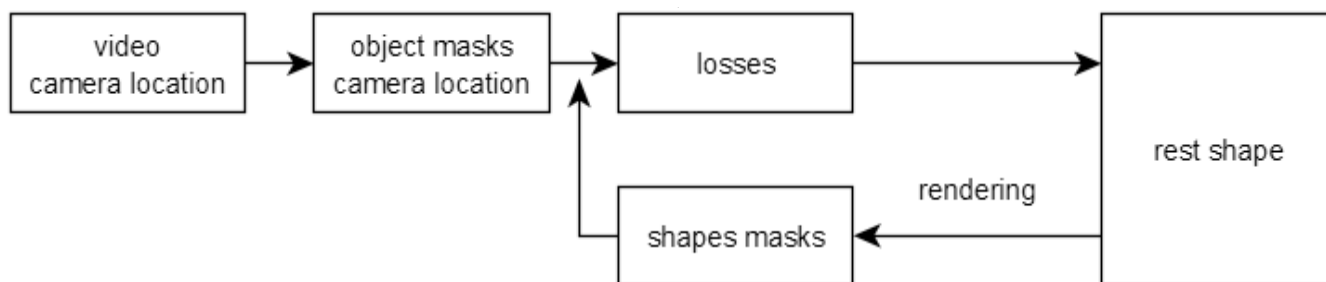


Рис. 1. Структурная схема процесса восстановления формы

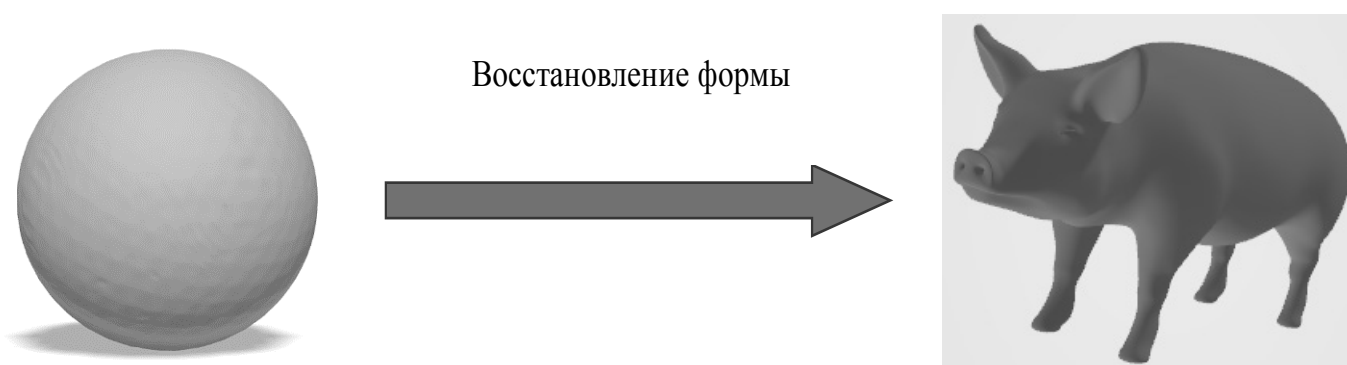


Рис. 2. Идея преобразования шара в искомый объект

На сегодняшний день существуют подходы для решения данной задачи. Так, например, СЗДРО — модель, которая способна реконструировать форму объекта по монокулярным представлениям. Но недостаток данного метода заключается в том, что помимо информации о ракурсах камеры необходимо предварительно расставить маркеры аннотации на 2Д объект. Алгоритм NRSfM, который учитывает набор траекторий особых точек на 2Д пространстве, так же способен восстанавливать 3Д форму [11, 12, 13]. Его проблема заключается в том, что метод очень требовательно относится к надежности таких траекторий, хоть и решает текущую задачу без заданной конкретной формы для определенной категории [14, 15, 16].

Недавний прогресс в дифференцируемом рендеринге позволяет переформулировать задачу как задачу анализа путем синтеза: обратная графическая задача восстановления формы 3D-объекта. Схема подхода изображена на рисунке 1.

На первом этапе производится подготовка данных, представляющих собой коллекцию изображений с целевым объектом и углы расположения камеры, снимающей этот объект. Используя претренированную модель, можно извлечь 2Д маски объекта, которые в свою очередь будут проекциями 3Д формы.

Далее необходимо настроить цикл оптимизации для подгонки сетки шаблона к наблюдаемым изображениям набора данных на основе потери визуализированного силуэта. За шаблон формы был выбран шар. Идея состоит в том, чтобы модель, имея на входе 2Д маски объекта, смогла преобразить шар в искомую форму (рис. 2).

Модель работает так же, как и любая другая модель машинного обучения, используя градиентный спуск и обновляя параметры модели. Алгоритму нужны только результаты видео и сегментации, чтобы учиться, преобразуя обратно визуализированный объект в сегментированное изображение и сравнивая его с входным сигналом. Что еще лучше, так это то, что все это делается в процессе итеративного самостоятельного обучения.

Для всех манипуляций использовались следующие функции потерь [17]:

1) *Edge-loss*. Регуляризатор позволит улучшить качество итоговой формы *mesh*, используя среднее значение длины ребер:

$$L_{edge}(V, E) = \frac{1}{|E|} \sum_{(v, v') \in E} \|v - v'\|^2, \quad E \subseteq V \times V,$$

где  $V$  — множество вершин полигональной модели, а  $E$  — множество ребер модели.



Рис. 3. Пример реальных объектов для реконструкции

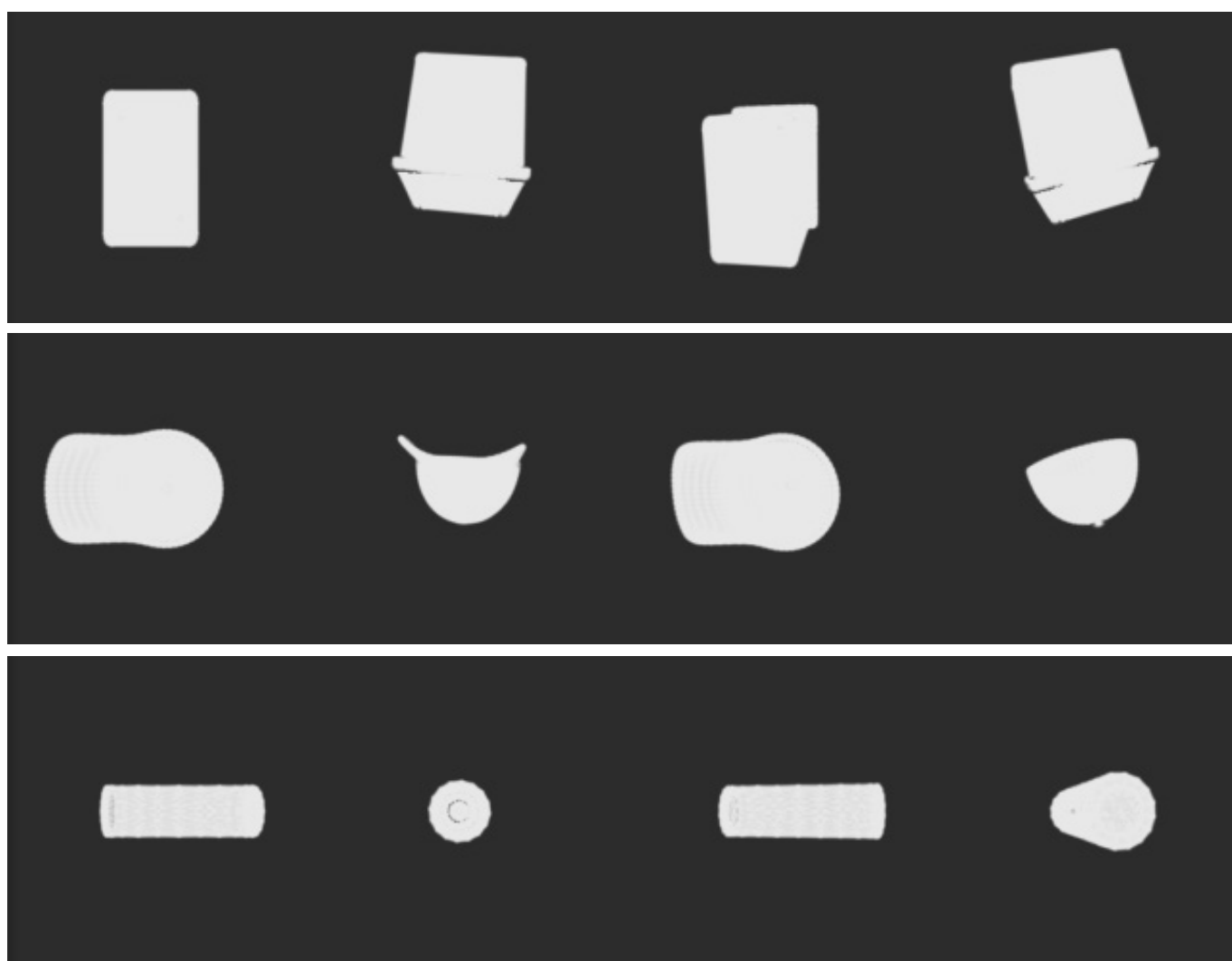
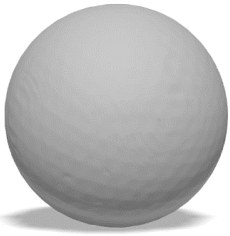
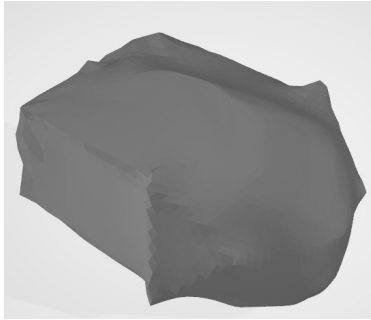
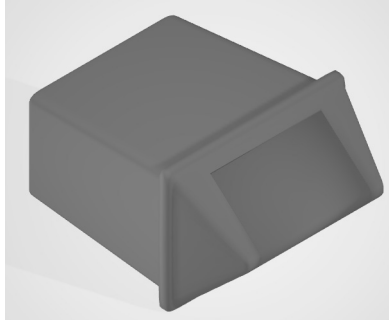
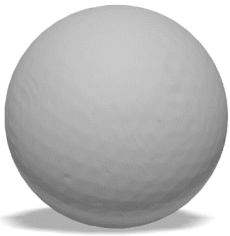
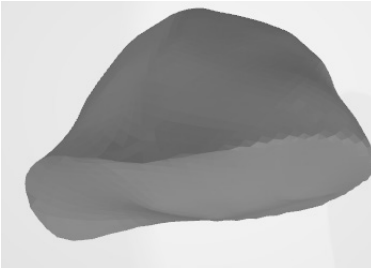
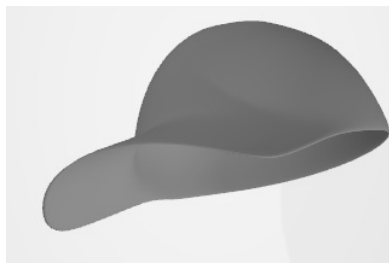
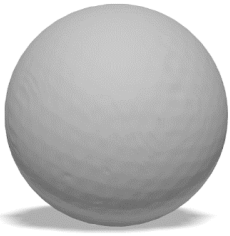
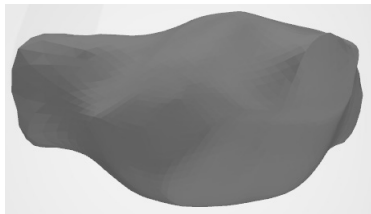



Рис. 4. Маски объектов из тестового набора данных

Таблица 1. Результаты восстановления после нескольких итераций.

Итерации	0	300	4500
Мусорное ведро			
Бейсболка			
Батарейка			

2) *Normal-loss*. Показывает, насколько сильно различаются поля нормалей у двух полигональных моделей, т.е., минимизируя данный критерий, происходит уменьшение углов между соответствующими нормальями:

$$L_{norm}(P, Q) = -|P|^{-1} \sum_{(p,q) \in \Lambda_{P,Q}} |u_q \cdot u_p| - |Q|^{-1} \sum_{(q,p) \in \Lambda_{Q,P}} |u_p \cdot u_q|,$$

где  $P$  и  $Q$  — два точечных множества,  $u_p$  и  $u_q$  — единичная нормаль к произвольным точкам  $p$  и  $q$  соответственно.

3) *Laplacian-loss*. Если оператор преобразования вершины *mash* в Лапласовы координаты будет иметь вид:

$$\delta_p = p - \sum_{k \in N(p)} \frac{1}{\|N(p)\|} k,$$

тогда Лапласов регуляризатор определяется как:

$$L_{Lap}(x) = \sum_p \left\| \delta'_p - \delta_p \right\|_2^2,$$

где суммирование производится по всем вершинам меша,  $N(p)$  — множество вершин, которые связаны с данным ребром, а штриховая Лапласова координата обозначает *mash* на предыдущей итерации.

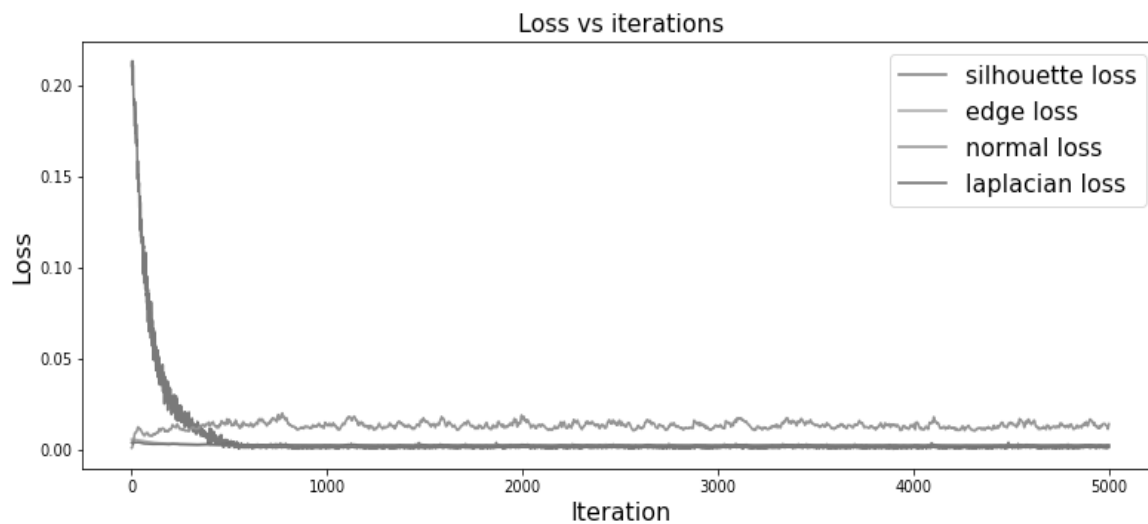
4) *Silhouette*. Средняя потеря силуэта как расстояние L2 между полученным и целевым силуэтами.

#### Проведение эксперимента

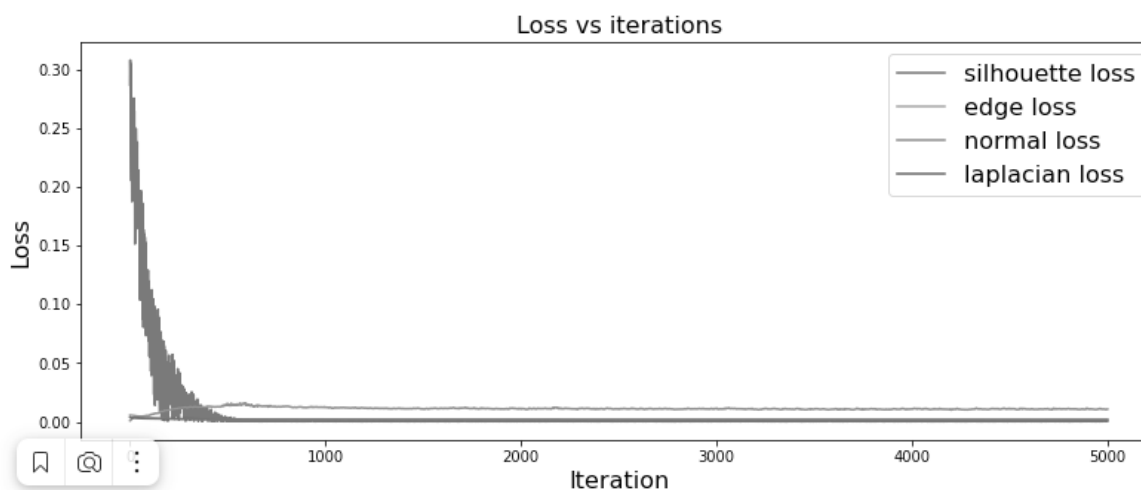
В качестве проверки работоспособности и конкурентоспособности алгоритма были выбраны некоторые предметы (рис. 3), по которым собран набор данных для тестирования.

Из коллекции изображений образованы маски объекта под разными углами, которые и использовались в качестве входных данных (рис. 4).

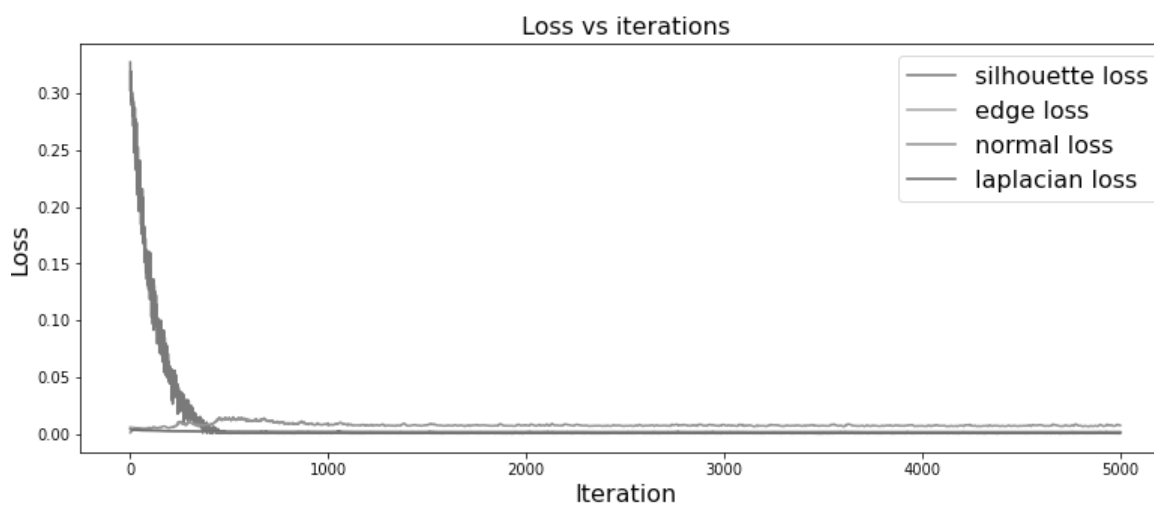
У такого подхода имеется ограничение, выявленное эмпирически: алгоритм с трудом оценивает



(a)



(б)



(в)

Рис. 1. Графики зависимостей функций потерь от итераций: а — мусорное ведро, б — бейсболка, в — батарейка.

поверхности, которые не видны ни в одном входном представлении, и терпит неудачу при сильных перекрытиях, которые пропускаются аннотациями маски. Его эффективность также нуждается в улучшении.

### Заключение

В работе рассмотрен подход восстановления 3D формы по монокулярному видео без шаблонов, специфичных для конкретной категории, что делает его применимым к широкому спектру сценариев.

Преимущество метода очевидно и заключается в отсутствие необходимости предварительно обучать модель на заранее подготовленных 3D данных из-за способности к итеративному обучению.

В качестве улучшений, в дальнейшем планируется модернизировать алгоритм для реконструкции нежестких форм, путем анализа оптического потока или поиска оптимальных точек сопряжения. И основной задачей является автоматическое определение положения камеры относительно объекта, чтобы стать абсолютно автономным способом реконструкции формы объекта.

### ЛИТЕРАТУРА

1. Armino Cachada. How to turn 2D photos into a 3D model using Nvidia Kaolin and PyTorch [Электронный ресурс], 2021. Режим доступа: <https://spltech.co.uk/how-to-turn-2d-photos-into-a-3d-model-using-nvidia-kaolin-and-pytorch-a-3d-deep-learning-tutorial/>
2. Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In CVPR, 2020
3. Angjoo Kanazawa, Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In ECCV, 2018.
4. Georgia Gkioxari, Jitendra Malik, and Justin Johnson. Mesh r-cnn. In ICCV, pages 9785–9795, 2019.
5. Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In CVPR, 2020
6. Angjoo Kanazawa, Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In ECCV, 2018.
7. Xueting Li, Sifei Liu, Kihwan Kim, Shalini De Mello, Varun Jampani, Ming-Hsuan Yang, and Jan Kautz. Self-supervised single-view 3d reconstruction via semantic consistency. ECCV, 2020.
8. Thomas J Cashman and Andrew W Fitzgibbon. What shape are dolphins? building 3d morphable models from 2d images. PAMI, 35(1):232–244, 2012.
9. Shubham Goel, Angjoo Kanazawa, and Jitendra Malik. Shape and viewpoints without keypoints. In ECCV, 2020
10. Christopher B Choy, JunYoung Gwak, Silvio Savarese, and Manmohan Chandraker. Universal correspondence network. In NeurIPS. 2016
11. Ignacio Rocco, Mircea Cimpoi, Relja Arandjelovic, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Neighbourhood consensus networks. In NeurIPS, 2018.
12. Christoph Bregler, Aaron Hertzmann, and Henning Biermann. Recovering non-rigid 3d shape from image streams. In CVPR, volume 2, pages 690–696. IEEE, 2000
13. Paulo FU Gotardo and Aleix M Martinez. Non-rigid structure from motion with complementary rank-3 spaces. In CVPR, pages 3065–3072. IEEE, 2011.
14. Chen Kong and Simon Lucey. Deep non-rigid structure from motion. In ICCV, pages 1558–1567, 2019.
15. Peter Sand and Seth Teller. Particle video: Long-range motion estimation using point trajectories. IJCV, 80(1):72, 2008.
16. Vikramjit Sidhu, Edgar Tretschk, Vladislav Golyanik, Antonio Agudo, and Christian Theobalt. Neural dense nonrigid structure from motion with latent space constraints. In ECCV, pages 204–222. Springer, 2020.
17. Narayanan Sundaram, Thomas Brox, and Kurt Keutzer. Dense point trajectories by gpu-accelerated large displacement optical flow. In ECCV, pages 438–451. Springer, 2010
18. 3D ML. Часть 2: метрики качества и функции потерь в задачах 3D ML [Электронный ресурс]. -Режим доступа: <https://medium.com/phygitalism/3d-ml-metrics-loss-functions-9708ff0476e2>

© Маньшин Илья Михайлович (manhin@yandex.ru),

Фальков Георгий Александрович (bag.falkova@gmail.com), Зуев Сергей Валентинович (sergey.zuev@bk.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»