

Разработка платформы для реализации алгоритма восстановления трехмерной модели объекта по однокурсным снимкам

П.В. Каткова¹, П.Ю. Якимов¹

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. На сегодняшний день область компьютерного зрения активно развивается, и вместе с ростом количества внедрений систем компьютерного зрения возрастает необходимость в качественных реализациях алгоритмов 3D-реконструкции. Среди типов 3D-реконструкции наибольшая потребность в использовании имеется у алгоритмов, которые, в качестве параметра, принимают один снимок. В данной статье приведен обзор наиболее популярных существующих алгоритмов восстановления 3D модели, а также описана предполагаемая реализация платформы для восстановления трехмерной модели объекта по однокурсным снимкам.

1. Введение

В настоящее время большое значения имеет использование 3D моделей вместо реальных физических объектов. Это объясняется тем, что, цифровые копии предоставляют пользователю большую вариативность и гибкость в процессе работы. Кроме того, использование цифровой модели существенно экономит время, за счет того, что ее можно использовать вне зависимости от местонахождения прототипа модели, а также за счет того, что процесс вычисления характеристик объекта происходит на несколько порядков быстрее, при использовании его цифровой модели[1].

Для восстановления трехмерных моделей по двумерным данным используются методы 3D-реконструкции. Значение задачи восстановления трехмерных объектов растет с каждым днем. Спектр ее применения охватывает такие области как медицина (например, в компьютерной томографии), компьютерное зрение (восстановление сцен, например для построения траектории движения), микроскопия, кинематография, мультипликация, видеонаблюдение (например, для биометрической идентификации человека) и т.д.

В случае, если входные состоят из набора изображений (пример процесса получения этого набора изображен на рисунке 1), то результат восстановления практически однозначен. Несколько лет назад, компания Autodesk выпустила продукт по названию Rescap, который может воссоздавать 3D модель по набору фотографий [2]. Однако, в реальном мире преобладает необходимость восстановления 3D модели по одному снимку, так как это бо́льшую практическую применимость.

Проблемой восстановления модели по одному снимку является неоднозначность определения фигуры с обратной стороны (невидимой на снимке) [3]. Такая же проблема относится к текстурированной, если с видимой на снимке стороны объекта ее можно частично

скопировать, то для обратной стороны модели должен быть разработан алгоритм составления текстуры.



Рисунок 1. Студия, позволяющая получить набор снимков с разных ракурсов [4].

2. Обзор существующих решений данной задачи

2.1 Алгоритмы восстановления фигуры и позы человека

Алгоритм End-to-end Recovery of Human Shape and Pose, представленный в июне 2019 года, позволяет восстанавливать модель человека с учетом его позы по единственной фотографии. В отличие от методов аналогов, данный алгоритм может определять местоположения ключевых точек даже когда человек на фотографии находится под углом.

На вход в алгоритм подается RGB изображение, которое сначала проходит через сверточный кодер. Далее, результат поступает в модуль 3-D регрессии, который итеративно минимизирует ошибку 3D модели. Полученный результат передается в модуль дискриминанта, который определяет является ли полученная модель моделью человека или нет. Схема работы данного алгоритма представлена на рисунке 2.

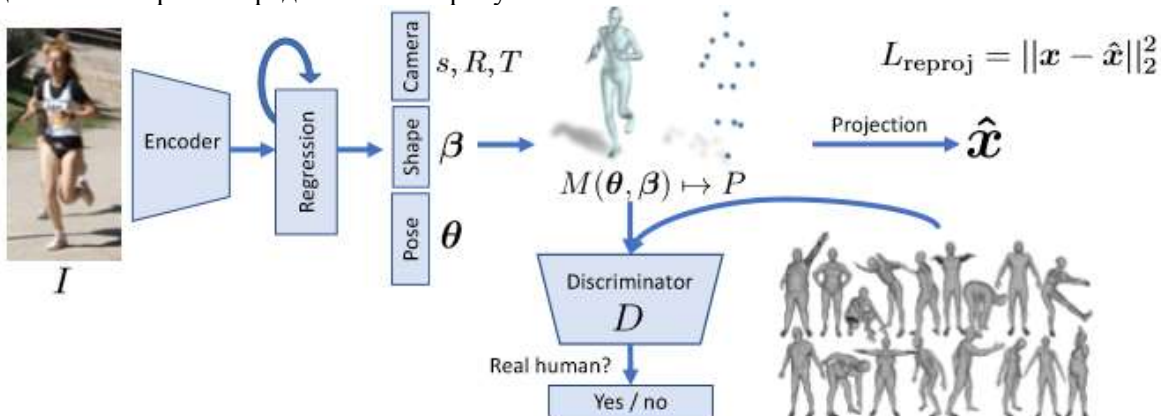


Рисунок 2. Схема работы алгоритма End-to-end Recovery of Human Shape and Pose [5].

Для данного метода был проведен ряд экспериментальных исследований. На рисунке 3 продемонстрировано сравнение потерь при 3D-реконструкции различных методов.

Ниже, на рисунке 4 представлено время восстановления для разных методов.

По рисункам 3 и 4 видно, что данный метод показывает наилучшие результаты по сравнению с другими представленными методами.

Method	Reconst. Error
Rogez <i>et al.</i> [35]	87.3
Pavlakos <i>et al.</i> [33]	51.9
Martinez <i>et al.</i> [26]	47.7
*Regression Forest from 91 kps [20]	93.9
*SMPLify [5]	82.3
*SMPLify from 91 kps [20]	80.7
*HMR	56.8
*HMR unpaired	66.5

Рисунок 3. Сравнение HMR с другими методами по критерию ошибки 3D-реконструкции [5].

Method	Fg vs Bg		Parts		Run Time
	Acc	F1	Acc	F1	
SMPLify oracle[20]	92.17	0.88	88.82	0.67	-
SMPLify [5]	91.89	0.88	87.71	0.64	~1 min
Decision Forests[20]	86.60	0.80	82.32	0.51	0.13 sec
HMR	91.67	0.87	87.12	0.60	0.04 sec
HMR unpaired	91.30	0.86	87.00	0.59	0.04 sec

Рисунок 4. Сравнение HMR с другими методами по критерию скорости 3D-реконструкции [5].

2.2 Алгоритмы восстановления модели человека

Предыдущий алгоритм имел функциональность восстанавливать фигуру человека, без учета формы одежды, а данный алгоритм восстанавливает модель человека с учетом предметов одежды, а также накладывает текстуру на полученный 3D объект. Данный метод был представлен в августе 2018 года.

Алгоритм работает следующим образом: первым шагом извлекается 2D силуэт и создается 3D карта с расположением суставов модели, далее происходит генерация новых 2D силуэтов модели с использованием полученной карты расположения суставов, после этого, по полученному набору 2D силуэтов производится 3D-реконструкция модели. Если 2D силуэты построены некорректно, то сетка для восстановления объекта также не будет соответствовать модели. Чтобы избежать данной ситуации, метод использует алгоритм глубокого распознавания поверхности, который включает в себя «жадную выборку». Последним шагом происходит нанесение текстуры на реконструированный объект.

Схема работы представленного выше алгоритма изображена на рисунке 3.

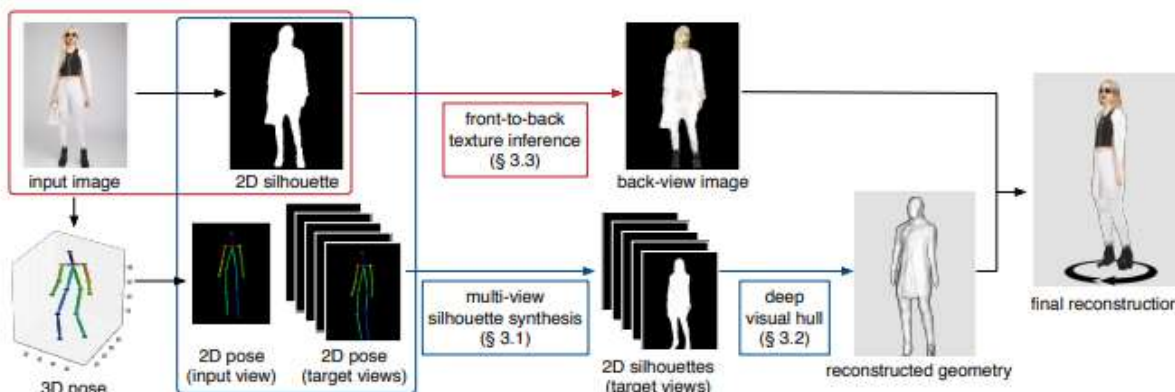


Рисунок 5. Схема работы алгоритма SiCloPe: Silhouette-Based Clothed People [6].

В ноябре 2019 года был опубликован метод PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization [7]. Данный метод позволяет восстанавливать 3D модель по одной или нескольким снимкам, а его особенностью является качественное построение текстуры даже со стороны объекта, невидимой на снимке.

Данный метод позволяет качественно восстанавливать даже сложные фигуры, содержащие в себе мятую одежду, высокий каблук или комплексную прическу.

Алгоритм состоит из сверточного кодера и непрерывной функции. Схема работы данного алгоритма представлена на рисунке 6.

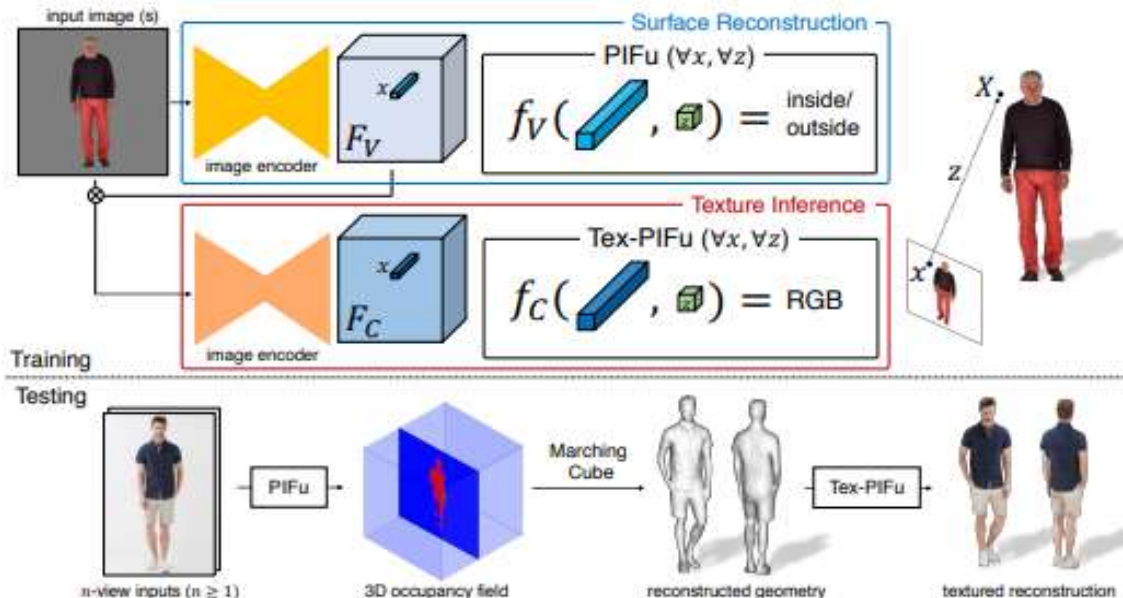


Рисунок 6. Схема работы алгоритма PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization [7].

2.3 Алгоритмы параметризованного восстановления человека

Алгоритм SMPL: A skinned multi-person linear model [8] является одним из наиболее популярных параметризованных методов воссоздания модели. Он был представлен в 2015 году и до сих пор используется в методах 3D восстановления, как для сравнения, так и для реализации части метода.

SMPL была обучена на нескольких тысячах 3D объектах человеческого тела, имеющих разную форму и фигуру. Результирующие модели тела человека имеют карту с весом в каждой точке его тела, таким образом, при смене поз, места сгибов или поворотов конечностей смотрятся наиболее естественно.

Полученные, в результате работы данного алгоритма, 3D модели могут использоваться в таких программах, как Autodesk Maya или Unity с последующим анимированием.

Ниже, на рисунке 6 проиллюстрирована модель SMPL.

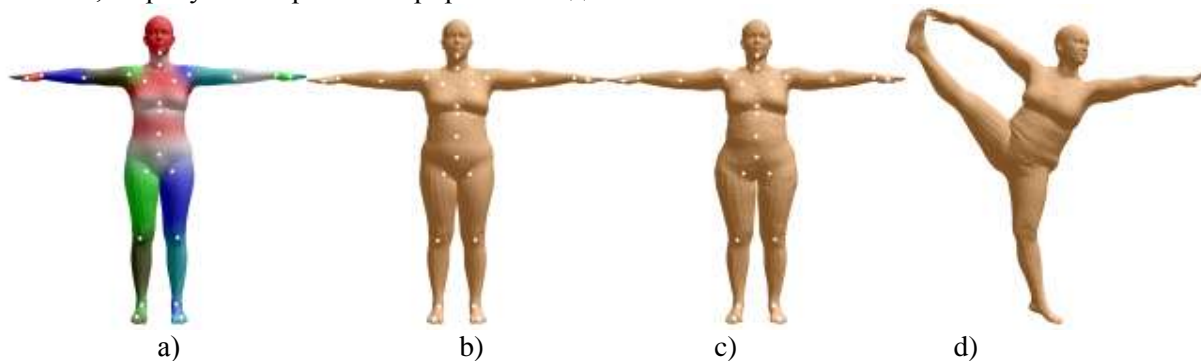


Рисунок 6. Модель SMPL: (a) – модель тела человека с сеткой весов, (b) – параметризованная модель тела человека, (c) – модель тела с формой тела, полученной в результате смешивания форм, (d) – модель тела человека с добавленной позой [8].

3. Экспериментальные исследования

В процессе эксперимента был протестирован алгоритм PIFu на различных данных. В качестве входных данных были использованы изображения и соответствующие им маски, созданные с использованием программы Photoshop. Изображения имеют формат png и разрешение 720x1080 пикселей. Исходные изображения и полученные данные для каждого из трех экспериментов проиллюстрированы на рисунках 7-9.

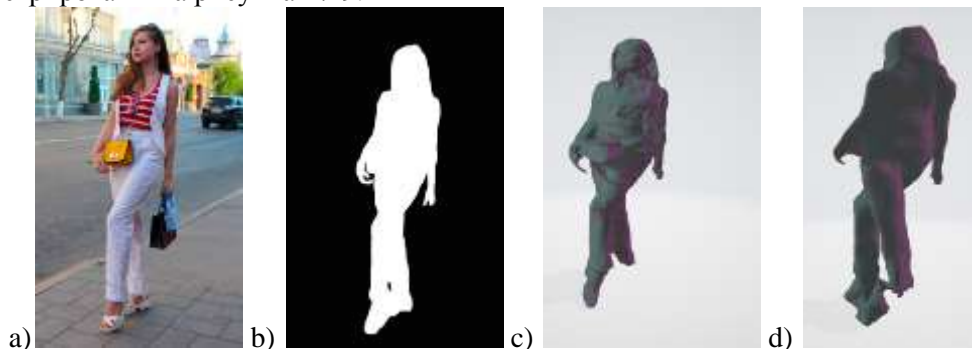


Рисунок 7. Данные первого эксперимента: (a) – исходное изображение, (b) – маска исходного изображения, (c) – вид полученного 3D объекта спереди, (d) – вид полученного 3D объекта сзади.

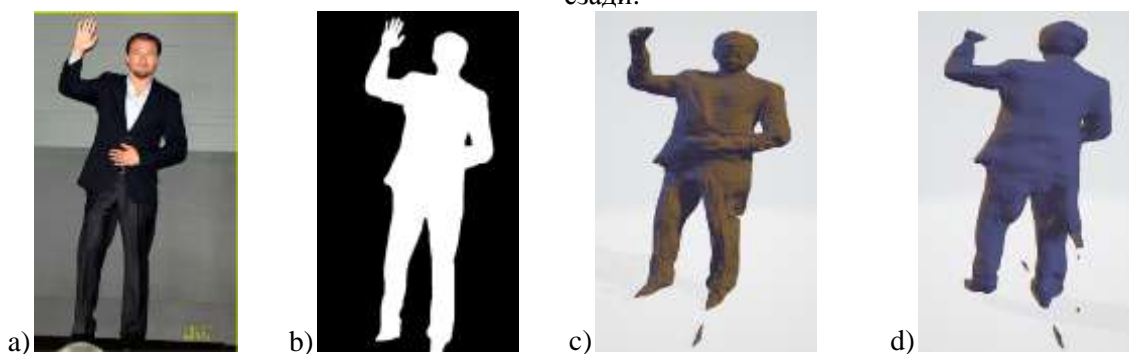


Рисунок 8. Данные второго эксперимента: (a) – исходное изображение, (b) – маска исходного изображения, (c) – вид полученного 3D объекта спереди, (d) – вид полученного 3D объекта сзади [9].

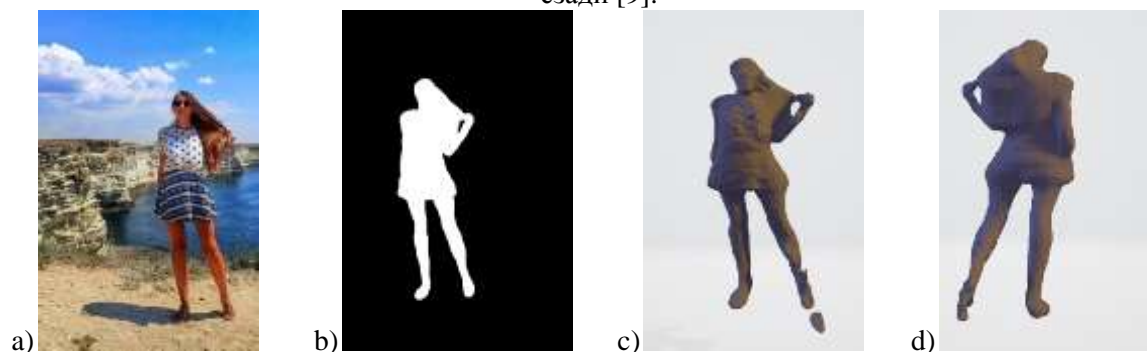


Рисунок 9. Данные третьего эксперимента: (a) – исходное изображение, (b) – маска исходного изображения, (c) – вид полученного 3D объекта спереди, (d) – вид полученного 3D объекта сзади.

Скорость работы алгоритма PIFu в первом случае составила 8.92 секунд, во втором - 10.47 секунд и в третьем - 7.19 секунд.

В данной работе предполагается реализовать алгоритм для воссоздания 3D модели человека по одной фотографии. Данный алгоритм, впоследствии, можно будет использовать для создания виртуальной примерочной. В качестве примера будут использованы, рассмотренные выше, методы восстановления фигуры человека.

Метод будет включать в себя использование двух сверточных нейронных сетей для восстановления модели и для восстановления текстуры.

В качестве платформы для пользователя предполагается сделать Telegram бот, который будет принимать изображение и по результату работы алгоритма возвращать 3D модель.

Для реализации планируется создать датасет, содержащий набор из 2D изображений людей и соответствующим им 3D объектов. Для создания 3D объектов для этого набора данных будут использованы камеры Microsoft Kinect 2.0 и stereocam ZED 2K. Камера Microsoft Kinect использует инфракрасный лазер для определения глубины матрицы изображения. Оптимальное расстояние для распознавания объектов у Kinect варьируется от 1 до 4 метров [10]. В отличие от Kinect ZED не имеет инфракрасного датчика, и для определения глубины применяются методы, использующие искусственный интеллект [11].

4. Заключение

Таким образом, были рассмотрены методы восстановления 3D модели человека, методы восстановления фигуры и позы человека по одному снимку, а также алгоритм SMPL, который воссоздает 3D модель человека по заданным параметрам. Также, для каждого метода был рассмотрен подход к разработке.

В результате обзора была поставлена предлагаемая реализация и было определено, что для решения задачи создания виртуальной примерочной необходимо реализовать метод восстановления фигуры человека.

5. Благодарности

Работа выполнена в рамках государственного задания по теме FSSS-2020-0017 при частичной поддержке РФФИ: проект № 17-29-03112 офи_м и проект № 19-29-01235 мк».

6. Литература

- [1] Евсеев, О.В. Разработка и исследование моделей и алгоритмов реконструкции трехмерной плотности облака точек, заданного серией параллельных сечений, 2016.
- [2] Autodesk ReCup // Autodesk Knowledge Network, 2019 [Electronic resource]. – Access mode: <https://help.autodesk.com/view/RECAP/2018/ENU/?guid=GUID-91810DA7-71F5-4C6C-989C-790455B7C847> (25.11.2019).
- [3] Choi, W. Understanding indoor scenes using 3d geometric phrases // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013. – P. 33-40.
- [4] Photogrammetry // Imgur, 2019 [Electronic resource]. – Access mode: <https://imgur.com/gallery/yuEncdf/comment> (09.11.2019).
- [5] Kanazawa, A. End-to-end recovery of human shape and pose // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018. – P. 7122-7131.
- [6] SiCloPe: Silhouette-Based Clothed People [Electronic resource]. – Access mode: <https://arxiv.org/pdf/1901.00049v2.pdf> (30.11.2019).
- [7] Saito, S. PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization // arXiv preprint arXiv: 1905.05172, 2019.
- [8] Loper, M. SMPL: A skinned multi-person linear model // ACM transactions on graphics (TOG). – 2015. – Vol. 34(6). – P. 248.
- [9] Leonardo DiCaprio Club, 2020 [Electronic resource]. – Access mode: <https://ru.fanpop.com/clubs/leonardo-dicaprio/images/10841990/title/leonardo-dicaprio-photo/> (10.02.2020).
- [10] Kinect for Windows SDK. Sensor, 2020 [Electronic resource]. – Access mode: <https://habr.com/ru/post/150955/> (01.02.2020).
- [11] ZED 2 // Stereolabs, 2020 [Electronic resource]. – Access mode: <https://www.stereolabs.com/zed-2/> (01.02.2020).

3D Reconstruction via single 2D Image

P.V. Katkova¹, P.Y. Yakimov¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. Computer Vision technology is rapidly developing nowadays. The need for 3D-reconstruction methods increases along with a number of Computer Vision system implementation. The highest need is for methods, which are using single image as an input data. This article provides an overview of existing methods for 3D-reconstruction and an explanation of planned implementation, which consists of a platform and a 3D-reconstruction algorithm using single image.