

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Барabanова Татьяна Александровна

Выпускная квалификационная работа бакалавра

**Построение трехмерной модели лица по набору
изображений**

Направление 01.03.02

«Фундаментальная информатика и информационные технологии»

ООП «Программирование и информационные технологии»

Научный руководитель:
кандидат физ.-мат. наук,
доцент
Погожев С. В.

Санкт-Петербург

2019

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
ПОСТАНОВКА ЗАДАЧИ.....	4
ГЛАВА 1. МЕТОДЫ ПОСТРОЕНИЯ ТРЁХМЕРНЫХ МОДЕЛЕЙ ПО НАБОРАМ ИЗОБРАЖЕНИЙ.....	5
1.1 Этапы реконструкции трёхмерных сцен.....	5
1.2 Отличительные особенности реконструкции трёхмерной модели лица.....	11
ГЛАВА 2. СОВРЕМЕННЫЕ ПОДХОДЫ К РЕКОНСТРУКЦИИ МОДЕЛИ ЛИЦА.....	15
2.1 Использование искусственных нейронных сетей.....	15
2.2 Трёхмерная морфируемая модель лица.....	17
ГЛАВА 3. РЕАЛИЗАЦИЯ АЛГОРИТМА.....	22
3.1 Язык и среда разработки.....	22
3.2 Используемые библиотеки.....	22
3.3 Описание алгоритма.....	24
3.4 Эксперименты.....	26
ЗАКЛЮЧЕНИЕ.....	29
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	30

ВВЕДЕНИЕ

Задача построения трехмерной модели лица является фундаментальной проблемой компьютерного зрения необычайной сложности. На данный момент эта задача является широко востребованной в таких сферах, как системы распознавания, визуализация трехмерных объектов, виртуальная реальность, компьютерная анимация. Существует большое количество работ, в которых предлагается множество различных подходов к решению данной проблемы, а также множество вариантов её решения. Но, несмотря на многолетние исследования, эта проблема все еще остается открытой.

Самым точным способом получения трёхмерной модели лица является, конечно же, использование специализированных устройств, таких как лазерные 3D сканеры. Но такие устройства имеют весьма внушительный ценник и требуют создания специализированного окружения, которое можно обеспечить только в лабораторных условиях, что значительно сужает область их применения, а также эти устройства не являются абсолютно безопасными для людей. Поэтому необходимо придумывать другие методы, более доступные, дешевые и безопасные.

Самым распространённым программным способом построения трёхмерной модели лица является его реконструкция на основе набора фотографий лица с разных ракурсов. В связи с тем, что данные системы предполагают наличие именно нескольких изображений лица в качестве входных данных, приходится решать ряд дополнительных задач, таких как нахождение зависимостей между положениями лица, разными выражениями и неравномерным освещением на разных фотографиях. В целом, эти методы требуют сложных и, зачастую, неэффективных решений. Поэтому в данной работе особое внимание будет уделено подходам к реконструкции трёхмерной модели лица с использованием только одной фотографии.

ПОСТАНОВКА ЗАДАЧИ

Целью данной работы является изучение общих подходов к трёхмерной реконструкции лиц на основе наборов изображений, изучение существующих современных тенденций и методов решения данной задачи. Также, в данной работе будет рассматриваться задача реконструкции трёхмерной модели лица на основе только одной фотографии. В конечном итоге необходимо реализовать алгоритм, выполняющий реконструкцию трёхмерной модели лица с использованием одного или нескольких изображений самостоятельно на основе уже существующих решений.

Постановка задачи для реализации алгоритма следующая. Пусть имеется одно или несколько изображений лица. Положение лица на фотографии, эмоция, изображаемая человеком на фото, могут быть произвольными, также условие, что фотография является полностью фронтальной, не является обязательным. В качестве входных данных также может использоваться набор заранее размеченных ключевых точек лица на фотографии. Необходимо наиболее точно построить трехмерную модель лица, изображенного на фотографии, и наложить на эту модель текстуру исходного лица.

ГЛАВА 1. МЕТОДЫ ПОСТРОЕНИЯ ТРЁХМЕРНЫХ МОДЕЛЕЙ ПО НАБОРАМ ИЗОБРАЖЕНИЙ

Задача реконструкции трёхмерной модели лица является частной задачей реконструкции трёхмерной сцены, где единственным объектом, представленным на сцене, является голова, а именно – лицо. В данном разделе будут рассмотрены общие подходы к решению задачи построения трёхмерной сцены по набору фотографий, а так же особенности, присущие реконструкции модели лица.

1.1 Этапы реконструкции трёхмерных сцен

Решение задачи построения трёхмерной модели сцены по набор фотографий можно разделить на четыре основных этапа [1]:

- 1) Производство, либо получение, изображений/фотографий сцены, подлежащей 3D-реконструкции;
- 2) Сопоставление расположения одних и тех же объектов сцены на разных изображениях;
- 3) Нахождение пространственного расположения объектов сцены;
- 4) Построение модели трёхмерной сцены по вычисленным координатам характеристических точек в трёхмерном пространстве.

Первый этап заключается в получении исходных данных для решения задачи. Для построения трёхмерной модели сцены обычно используются несколько фотографий этой сцены, сделанные с разных ракурсов. Чаще всего используют только два изображения сцены, представляющие сцену с немного различающихся ракурсов. Такая пара плоских изображений называется стереопарой[2].

Второй этап заключается в сравнении изображений и сопоставлении положения соответствующих объектов на изображениях. То есть на этом этапе осуществляется поиск точек на нескольких входных изображениях, соответствующих одним и тем же областям исходной трёхмерной сцены. Алгоритмы поиска этих точек исходят из предположения, что имеющиеся фотографии сцены, сделанные с разных ракурсов, были получены при одинаковых условиях (например, при одинаковом освещении), а также имеют одинаковые характеристики изображения (яркость, контрастность, резкость и т.п.). Из этого условия следует, что описание окрестности некоторой точки на одном изображении должно примерно соответствовать описанию окрестности этой же точки на другом изображении.

Один из методов поиска таких точек заключается в нахождении характерных элементов сцены на исходных изображениях. Под такими элементами сцены обычно подразумевают угловые точки объектов, либо их границы (линии), либо же целые области изображения. На нескольких фотографиях необходимо обнаружить одни и те же характерные элементы, принадлежащие одному и тому же объекту, и соотнести их между собой. Существует множество алгоритмов обнаружения характерных элементов. Рассмотрим основные принципы алгоритмов нахождения угловых точек.

Самый простой алгоритм нахождения угловых точек объекта – алгоритм Моравеца [3]. В нём проводится анализ изменения интенсивности пикселей в квадратной области, центром которой является интересующий нас пиксель, при сдвиге данной области по 8 направлениям на один пиксель – вверх, вниз, вправо, влево и в четырёх направлениях по диагоналям. Стандартно берут область размером 3x3, 5x5 или 9x9 пикселей. Алгоритм состоит из четырёх шагов:

- 1) Вычисление изменения яркости для каждого направления сдвига квадратной области для каждого пикселя изображения по следующей формуле:

$$D_{a,b}(x, y) = \sum_{\substack{\forall u,v: \\ x+u, y+v \in h}} (I(x + a + u, y + b + v) - I(x + u, y + v))^2$$

где (x, y) – координаты рассматриваемого пикселя; (a, b) – направления сдвига относительно осей x и y соответственно, $(a, b) \in \{(0,1), (1,0), (1,1), (0, -1), (-1,0), (-1, -1), (1, -1), (-1,1)\}$; h – рассматриваемая квадратная область; $I(x, y)$ – яркость пикселя с координатами (x, y) ;

- 2) Угловая точка имеет смежные рёбра, поэтому необходимо найти, при сдвиге в каком направлении значение яркости будет минимальным. Это направление нам запоминать не обязательно, важно найти только угловую точку, а не определить вид угла. Поэтому просто из всех возможных направлений сдвига выбираем наименьшее значение яркости:

$$M(x, y) = \min(D_{a,b}(x, y)).$$

Значения, полученные с помощью данной формулы, называются картой вероятности нахождения углов в пикселях изображения;

- 3) Один из параметров алгоритма – пороговое значение яркости T . На данном этапе просто исключаем из рассмотрения пиксели, значения $M(x, y)$ в которых ниже порога T ;
- 4) Последний шаг заключается в исключении найденных повторяющихся угловых точек. Для этого используют алгоритм NMS [4], заключающийся в поиске локальных максимумов функции отклика. Оставшиеся ненулевые элементы карты вероятности нахождения углов $M(x, y)$ соответствуют искомыми угловым точкам (x, y) .

Данный метод достаточно прост в реализации, но он имеет значительные недостатки. Например, реальные фотографии обычно имеют значительный шум, метод Моравеца даёт много лишних срабатываний на границах объектов. Также этот метод плохо справляется с диагональными

рёбрами и не является независимым относительно преобразования изображения типа «поворот». В связи с данными недостатками метод Моравеца на практике почти не применяется, используют улучшенные модификации данного алгоритма.

Один из примеров таких модификаций – алгоритм нахождения угловых точек Харриса [5]. Данный алгоритм избавлен от одного из главных недостатков описанного ранее метода – отсутствие инвариантности к поворотам. Отличием от алгоритма Моравеца является то, что рассматриваются производные по некоторым значимым направлениям, функция яркости раскладывается в ряд Тейлора. Но данный алгоритм также плохо справляется с шумами на границах объектов.

На практике часто используют Scale-Invariant Feature Transform (SIFT) [6] и Speed Up Robust Feature (SURF) [7] – методы, которые заключаются в поиске локальных особенностей на изображениях и описания этих особенностей через дескриптор, для поиска точечных соответствий производится сравнение дескрипторов.

Помимо упомянутых алгоритмов, существует ещё множество других, но эти методы нахождения угловых точек не подходят для поиска особых точек на лице. Так как темой данной исследовательской работы является задача реконструкции лица, а не обычных трёхмерных сцен, я не буду рассматривать подробно данные алгоритмы. Для нахождения ключевых точек лица применяются другие методы, которые будут приведены далее в следующем разделе.

Третий этап заключается в нахождении пространственного расположения объектов сцены. Помимо координат особых точек, найденных на предыдущем этапе на изображениях, нам понадобятся параметры камер, с которых совершалась съемка рассматриваемых фотографий. Рассмотрим случай со стереопарой. Допустим, параметры камер представлены матрицами – P и P' . Координаты некой точки на двух изображениях, соответствующих камерам с параметрами P и P' – x и x' соответственно.

Тогда, пространственные координаты этой точки X можно вычислить, решив следующую систему уравнений:

$$\begin{cases} PX = x; \\ P'X = x'. \end{cases}$$

Методы, применяемые для решения данной задачи, глобально различаются способами калибровки камеры. Калибровка камеры – это задача получения внешних и внутренних характеристик камеры по имеющимся изображениям, полученным посредством этой камеры.

К характеристикам внутренней калибровки относят пять параметров [8] – два параметра α_x, α_y , соответствующие коэффициентам сжатия пикселей по ширине и по высоте изображения соответственно; два параметра u_0, v_0 , соответствующие координатам точки пересечения оптической оси и плоскости изображения; параметр $\gamma = \alpha_y * \text{tg}(\varphi)$ – коэффициент асимметрии между осями координат изображения, где φ – угол отклонения взаимного расположения осей координат изображения от стандартного, равного 90° . Эти параметры представляются в виде матрицы внутренней

калибровки: $K = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$. Параметры внутренней калибровки

определяются используемой камерой. Они будут одинаковы для всех входных фотографий, очевидно, при условии, что они сделаны с одной камеры.

К внешним параметрам калибровки камеры относят всего два параметра (R, T) , которые определяют преобразования координат, трансформирующие координаты трёхмерной сцены в систему координат камеры из мировой системы координат. R – 3×3 матрица вращения (либо 1×3 вектор вращения), T – 1×3 вектор сдвига. Внешние параметры для каждой фотографии будут индивидуальными.

Выделяют три группы алгоритмов:

- 1) Без калибровки камеры;
- 2) С калибровкой камеры;

3) С автокалибровкой.

Алгоритмы нахождения трёхмерных координат объектов восстанавливаемой сцены, не использующие предварительную калибровку камеры, основаны в основном на методах факторизации матриц.

Алгоритмы, требующие предварительную калибровку камеры, обычно применяются в случаях с небольшим количеством входных изображений сцены, например, при реконструкции по стереопаре. Классический алгоритм решения задачи калибровки камеры – алгоритм Tsai [9]. Этот алгоритм состоит из двух шагов, на первом определяются внешние характеристики калибровки камеры, а на втором – внутренние. Некоторые подходы к калибровке камеры основаны на наличии на сцене некоего калибровочного объекта, координаты которого известны в трёхмерном пространстве с большой точностью. Обычно такой объект состоит из двух или трёх ортогональных друг другу плоскостей. Пример такого алгоритма – технология, предложенная Zhengyou Zhang [10], в качестве калибровочного объекта в данном алгоритме выступает шахматная доска.

Класс алгоритмов, относящийся к группе с автокалибровкой камеры, не требует нахождения на сцене никаких калибровочных объектов. В данном подходе используется только информация с входных изображений, а именно особые точки и их соответствия между изображениями, найденные на предыдущем этапе. Обладая наборами соответствующих точек объектов на разных изображениях, можно легко восстановить положение камер в пространстве и относительно друг друга. И далее, по имеющимся данным, восстановить пространственные координаты объектов на сцене. Если изображения были сделаны одной и той же камерой, то внутренние характеристики калибровки будут константными. Пример такого алгоритма – алгоритм Bundle Adjustment [11].

Четвёртый этап заключается в построении 3D-модели сцены по найденным на предыдущем этапе пространственным координатам точек. Создается полигональная сетка, описывающая формы объектов.

Полигональная сетка представляет поверхность, состоящую из граней, рёбер и вершин, задающую форму объекта. Чем больше граней в создаваемой полигональной сетке, тем детальнее можно описать объект. В качестве граней полигональной сетки чаще всего используют треугольники, также гранями полигональной сетки могут являться четырёхугольники, либо простые выпуклые многоугольники большей размерности, если это необходимо. Такое деление поверхности на полигоны нужно для упрощения операций производимых над трёхмерными моделями. Чтобы создать видимость гладкой непрерывной поверхности необходимо использовать полигоны меньшей размерности и, соответственно, увеличивать число полигонов. Тогда, после наложения текстур, которое также осуществляется на данном этапе, полигональное деление будет практически незаметно.

1.2 Отличительные особенности реконструкции трёхмерной модели лица

Первый этап, получение фотографий, идентичен для любых реконструируемых сцен, в том числе и для сцены, на которой представлено лицо. Хочу отметить, в современных алгоритмах присутствует тенденция использования только одного изображения лица – фронтального. Подробнее этот аспект будет рассмотрен в следующей главе.

Самой главной отличительной чертой алгоритмов реконструкции трёхмерной модели лица по набору изображений от общего подхода к реконструкции трёхмерной сцены является второй шаг, а именно нахождение ключевых точек на лице. Первым этапом решения данной задачи является обнаружение лица на фотографии. Анализируя существующие алгоритмы, можно выделить четыре основных подхода [12]:

- 1) Эмпирические;
- 2) Поиск отдельных черт лица;

- 3) Использование шаблонов;
- 4) Использование обучаемых систем.

Эмпирические методы (Knowledge-Based Methods [13]) основаны на нашем представлении о лице. Алгоритмы пытаются описать человеческую логику, как человек, смотря на изображение, понимает, изображено ли там человеческое лицо, либо же это что-то иное. Программистом описываются определённые правила, которым должен соответствовать участок изображения, чтобы считаться изображением лица. К таким правилам можно отнести, например, то, что на лице должны присутствовать элементы, резко отличающиеся по интенсивности относительно основной части лица – рот, нос и два симметрично расположенных глаза, а основная часть лица, помимо этих элементов, должна быть достаточно однородной по цвету и яркости. К таким алгоритмам можно отнести метод Yang & Huang [14], который заключается в сильном сжатии изображения для сглаживания шумов. На маленьком изображении легче найти зону однородной яркости, то есть зону, где предпологаемо есть лицо. И после, в данной зоне произвести поиск сильно отличающихся по яркости зон, соответствующих глазам, носу и рту. Ещё одним алгоритмом, относящимся к группе эмпирических, является алгоритм Kotropoulos & Pitas [15]. В данном методе строятся вертикальная и горизонтальная гистограммы яркости:

$$HI(x) = \sum_{y=1}^n I(x, y), VI(y) = \sum_{x=1}^m I(x, y),$$

где $I(x, y)$ – яркость пикселя с координатами (x, y) .

Области с лицом, согласно данному алгоритму, должна соответствовать область изображения с непрерывно большими значениями на горизонтальной гистограмме яркости и с большими, но прерывающимися в местах нахождения глаз, носа и рта, значениями яркости на вертикальной гистограмме.

Второй подход поиска лиц на изображениях заключается в поиске характерных инвариантных признаков, инвариантность заключается в независимости от угла наклона и положения лица на изображении. В данных методах сначала осуществляется поиск характерных черт лица – глаза, нос и рот. Далее производится определение границ лица и объединение найденных признаков. Примером данного подхода является метод поиска лиц, предложенный Leung, Burl & Perona [16]. Этот алгоритм заключается в объединении локальных детекторов характерных черт лица со статистической моделью взаимного расположения этих черт на лице. Также Ещё один алгоритм с данным подходом – метод, описанный Yow & Cipolla [17]. Данный алгоритм производит поиск характерных точек на изображении, используя вторую производную гауссовского фильтра. Далее осуществляется поиск краёв с помощью порогового фильтра и производится группировку найденных черт в вероятные лица. Выбор наиболее вероятных черт лица осуществляется с помощью байесовской сети.

Использование шаблонов заключается в поиске соответствия заданному шаблону перебором всех возможных областей изображения. Например, метод трёхмерных форм, предложенный Sinha [18], использует шаблон, описываемый попарное отношение яркости различных областей на лице. Другой шаблон – модель распределения точек (PDM) [19]. Это статистическая модель, описывающая среднестатистическое лицо, представленное набором объектов, которые можно деформировать.

Четвертый подход заключается в использовании обучаемых систем. Данную задачу можно отнести к задаче классификации, при которой необходимо отнести рассматриваемую область изображения к одному из двух классов – лицо или не лицо. Алгоритмы строят для областей изображения вектора признаков, и на основании данных обучения принимается решение, к какому классу отнести рассматриваемую область. Обучение проводится на заранее подготовленном множестве изображений, относящихся к обоим классам. На данный момент этот подход является

самым популярным, так как алгоритмы дают очень хорошие результаты при использовании минимума ресурсов. Самый ресурсоёмкий этап – этап обучения системы, необходимо проводить всего один раз. Далее результаты обучения можно использовать для сколь угодно большого числа фотографий. К методам, использующим данный подход, относят столь популярный на данный момент метод искусственных нейронных сетей, также метод главных компонент, метод факторного анализа, метод опорных векторов и др.

Зачастую на практике вообще не используются алгоритмы для нахождения особых точек на фотографиях в задаче реконструкции трёхмерной модели лица. В таких случаях список необходимых ключевых точек подаётся на вход программы пользователем вместе с изображениями, либо в виде подготовленного пользователем заранее списка координат, либо же в программе реализован функционал, с помощью которого пользователь после загрузки фотографий самостоятельно графически отмечает требующиеся ключевые точки на загруженных изображениях.

ГЛАВА 2. СОВРЕМЕННЫЕ ПОДХОДЫ К РЕКОНСТРУКЦИИ ТРЁХМЕРНОЙ МОДЕЛИ ЛИЦА

В данной главе будет представлен обзор литературы, рассмотрены существующие современные подходы и выделены современные тенденции в решении задачи реконструкции лица по фотографиям. Одним из основных трендов современных алгоритмов является то, что в качестве входных данных алгоритмами используется всего одна фотография лица, причем, в ряде алгоритмов это изображение даже не обязательно должно быть фронтальным. Но, нужно понимать, что если какая-то часть лица будет не видна, то, например, будут пробелы в накладываемой текстуре. Или же, если алгоритмом предусмотрено «додумывание» невидимых частей, то результат может не всегда совпадать с действительностью, так как все реальные человеческие лица обладают особенностями, которые невозможно предсказать алгоритмически, так как они индивидуальны для каждого человека.

2.1 Использование искусственных нейронных сетей

Искусственные нейронные сети обретают всё большую популярность, эту технологию используют для решения совершенно разных задач в разных областях, и область трёхмерного моделирования также активно использует данную концепцию. Искусственная нейронная сеть – это математическая модель, приближенно описывающая работу и принципы организации биологической нейронной сети. Эта технология – попытка смоделировать и изучить процессы, протекающие в человеческом мозге для создания машин, обладающих искусственным интеллектом. Структура нейронной сети представляет собой совокупность искусственных нейронов – простых

процессоров, соединённых и взаимодействующих между собой [20].

Авторы работы [21] предлагают использовать свёрточную нейронную сеть (convolutional neural network, CNN) для решения задачи реконструкции лица из одного изображения. Свёрточные нейронные сети – это особая архитектура искусственных нейронных сетей, разработанная для наиболее эффективного решения задачи распознавания образов [22]. Идея структуры CNN основана на особенностях работы зрительной коры головного мозга – простые клетки сначала реагируют на прямые линии, по-разному расположенные, а реакция сложных клеток зависит от набора реакций простых клеток [23]. То есть восприятие нами изображения происходит от восприятия некоторых деталей к восприятию всей картины в целом. В упомянутой работе используют глубокое обучение нейронной сети на наборе данных [24] из пар 2D фотографий лиц людей, изображающих различные эмоции и расположенные под произвольным ракурсом, и соответствующих этим лицам 3D моделей.

В работе [25] предлагается использовать сразу три нейронные сети, одну кодирующую и две декодирующие. Кодирующая нейросеть принимает на вход одно изображение и извлекает из него текстуру лица, характеристики диффузной отражательной способности поверхности и параметры освещения. Далее, используя эти закодированные параметры в векторной форме, два декодера, в качестве которых используются свёрточные нейронные сети, переносят эти характеристики лица в трёхмерное пространство, определяя расположение бликов, диффузию и 3D текстуру лица. На основе полученных параметров, дифференцируемый рендеринг-слой генерирует модель лица посредством совмещения 3D текстуры, диффузии, освещения и параметров расположения камеры, полученных декодером. Этот способ показывает наилучшие результаты в сравнении с остальными, уже существующими.

2.2 Трёхмерная морфируемая модель лица

Один из самых популярных подходов к построению трёхмерной модели лица – использование трёхмерной морфируемой модели [26, 27] (3D Morphable Model, 3DMM). Эта модель была разработана достаточно давно, но пользуется большой популярностью в современных алгоритмах реконструкции лиц по фотографиям.

Суть трёхмерной морфируемой модели заключается в том, что изначально мы располагаем неким набором базисных лиц, полученных, к примеру, лазерным сканированием, на основе которых получаем усреднённую модель, представленную в векторном пространстве, которую можно изменять, морфировать. Математически это представляется следующим образом. Трёхмерную модель лица можно представить в виде вектора в $3 \cdot n$ -мерном пространстве, где каждая точка, описывающая модель, представлена тремя координатами $(x_i, y_i, z_i), i = \overline{1, n}$:

$$\vec{S} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ x_2 \\ \vdots \end{pmatrix} \in \mathbb{R}^{3 \cdot n}.$$

Текстуру лица также можно представить в виде в виде вектора в $3 \cdot n$ -мерном пространстве, где цвет каждой точки представлен в RGB цветовой модели и, соответственно, для однозначного определения цвета необходимо также три параметра, как и в случае с пространственными координатами, $(R_i, G_i, B_i), i = \overline{1, n}$:

$$\vec{T} = \begin{pmatrix} R_1 \\ G_1 \\ B_1 \\ R_2 \\ \vdots \end{pmatrix} \in \mathbb{R}^{3 \cdot n}.$$

Тогда сформированная база представляется набором пар векторов, советующих моделям отсканированных лиц и их текстурам: $\{\vec{S}_i, \vec{T}_i\}, i = \overline{1, N}$. Новые лица моделируются путем формирования линейных комбинаций векторов, представляющих базисные лица, с некими коэффициентами α_i , а текстуры, соответственно, формируются с помощью линейных комбинаций векторов текстур с параметрами β_i :

$$\vec{S} = \sum_{i=1}^N \alpha_i \cdot \vec{S}_i, \vec{T} = \sum_{i=1}^N \beta_i \cdot \vec{T}_i, \alpha_i, \beta_i \in [0,1].$$

Чтобы сделать модель ещё более вариативной можно представлять лицо не как один общий вектор, а разделить лицо на зоны. Отдельными векторами представлять такие зоны как лоб, глаза, нос, щёки, губы, подбородок. Тогда можно будет отдельно подбирать каждую черту лица, а не опираться на похожесть лица прототипа в целом.

Но для создания 3DMM обычно используется весьма ограниченное количество исходных лиц, около ста, к тому же, в основном берутся лица людей европеоидной расы. В связи с этим могут возникать проблемы в моделировании лиц людей разного возраста и разных рас.

К тому же, очевидно, что параметры α_i и β_i абсолютно любыми брать нельзя, так как будут получаться совершенно неестественные, искаженные лица, поэтому логично построить вероятностную модель для данных параметров, чтобы совсем неестественные результаты получались с очень низкой вероятностью. Для решения этой задачи к базисному множеству лиц применяют метод главных компонент, в качестве результата получаем усреднённое лицо $\vec{S}_0 = \frac{1}{N} \sum_{i=1}^N \vec{S}_i \in \mathbb{R}^{3 \cdot n}$, матрицу главных компонент $V \in \mathbb{R}^{3 \cdot n \times (N-1)}$ и множество вариаций вдоль главных компонент $\{\sigma_i^2\}_{i=1}^N$. Для генерации текстур аналогично. Тогда формирование нового лица можно представить как параметрическое изменение усреднённого лица, причем, предполагается, что вектор \vec{a} имеет Гауссовское распределение:

$$\vec{S} = \vec{S}_0 + V \cdot \vec{a}.$$

Вектор $\vec{\alpha}$ можно разделить на две составляющие, пусть одна отвечает за формы лица $\vec{\alpha}_{id}$, а другая за выражаемую лицом эмоцию $\vec{\alpha}_{exp}$, тогда получим:

$$\vec{S} = \vec{S}_0 + V_{id} \cdot \vec{\alpha}_{id} + V_{exp} \cdot \vec{\alpha}_{exp}.$$

В работе [28] используется именно этот подход. Используется не весь набор из множества базисных лиц, а только одна трехмерная усреднённая «эталонная» модель лица. Но так как средняя модель строилась, опять же, по ограниченному набору отсканированных лиц, этот подход не решает проблемы, связанной с реконструкцией лиц людей разных возрастов и рас.

Теперь необходимо рассмотреть обратную задачу, а именно задачу реконструкции лица с помощью трёхмерной морфируемой модели. Необходимо подобрать такие параметры для морфируемой модели, чтобы трёхмерная модель как можно более точно описывала лицо, представленное на входной фотографии. То есть необходимо минимизировать разность значений каждого пикселя на имеющейся фотографии (I_{target}) и сгенерированной ($I_{rendered}$). Причём, надо учитывать, что не все пиксели лица, присутствующие в трёхмерной модели будут видны на фотографии, например, нос может перекрывать часть щеки, если фотография не фронтальная. Для этого используется маска видимости \mathcal{M} . Математически это можно описать следующим образом:

$$E_{pix} = \frac{\|\mathcal{M} \cdot (I_{target} - I_{rendered})\|}{|\mathcal{M}|} \rightarrow min.$$

Но, помимо минимизации этого функционала, надо также учитывать ещё некоторые параметры. Например, у данной задачи могут существовать локальные минимумы, поэтому, чтобы их избежать, необходимо провести процесс инициализации. Инициализация заключается в сопоставлении ключевых точек на фотографии и на морфируемой трёхмерной модели.

Обычно используют 68 ключевых точек, расположенных следующим образом:

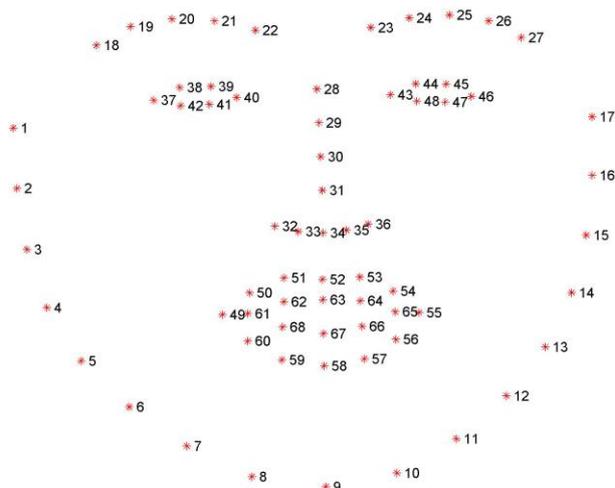


Рис. 1: Расположение 68 ключевых точек на лице.

Математически это может быть представлено минимизацией квадрата евклидового расстояния между двумя векторами, соответствующими ключевым точкам лица (L_{image}) и ключевым точкам построенной модели (L_{model}):

$$E_{lands} = \|L_{image} - L_{model}\|^2 \rightarrow \min.$$

Ещё необходимо помнить, что параметры трёхмерной морфируемой модели могут быть не абсолютно любыми, так как можно получить неестественные лица, эта проблема уже была рассмотрена ранее. Это ограничение параметров называется регуляризацией. Математически необходимо минимизировать следующую функцию:

$$E_{reg} = \sum_{i=1}^N \frac{\alpha_i^2}{\sigma_i^2} \rightarrow \min.$$

Собирая все эти параметры в общую формулу, получаем, что для решения задачи реконструкции трёхмерной модели лица с помощью морфируемой модели необходимо минимизировать следующую функцию:

$$E(\vec{\alpha}_{id}, \vec{\alpha}_{exp}, \vec{\beta}, \vec{p}) = E_{pix} + c_{lands} \cdot E_{lands} + c_{reg} \cdot E_{reg} \rightarrow min,$$

где вектор \vec{p} представляет некоторые прочие параметры, которые не были формализованы, например освещение, угол наклона головы и т.д. c_{lands} и c_{reg} – коэффициенты, которые подбираются вручную при реализации программы.

Решение задачи реконструкции лица по фотографии является более трудоёмким и ресурсоёмким процессом, чем простая генерация произвольных лиц на основе трёхмерной морфируемой модели. Но данный метод даёт действительно хорошие результаты. Главной сложностью является начальный этап создания базы трёхмерных моделей лиц, необходимо отсканировать очень большое количество людей разных рас, возрастов, чтобы создать модель, способную к трансформации в действительно любое лицо. Сейчас, так как трёхмерная морфируемая модель является достаточно популярным подходом в сфере 3D моделирования лиц, имеется ряд уже подготовленных баз лиц и даже уже готовых 3DMM.

ГЛАВА 3. РЕАЛИЗАЦИЯ АЛГОРИТМА

В данной главе будет представлено описание реализации программы, выполняющей реконструкцию трёхмерной модели лица по одной фотографии. Меня заинтересовал подход с использованием трёхмерной морфируемой модели (3DMM), эту технологию я и буду использовать.

3.1 Язык и среда разработки

Для реализации алгоритма был выбран язык программирования C++, так как на этом языке реализована одна из основных библиотек для работы с компьютерным зрением – OpenCV. Эта библиотека реализована и для ряда других языков программирования, но лучше всего с ней работать на языках C++, либо Python.

C++, вообще говоря, является традиционным языком для решения задач компьютерного зрения, программы, написанные на C++ весьма производительны и быстродейственны. Также легко реализуется кроссплатформенность. Я имею большой опыт работы с языком C++, нежели с Python, поэтому выбор был сделан в пользу первого.

Средой разработки была выбрана программный продукт компании Microsoft – Visual Studio 2017 [29].

3.2 Используемые библиотеки

В данной работе мной был использован ряд библиотек для работы с компьютерным зрением, с линейной алгеброй и для использования трёхмерной морфируемой модели: OpenCV, Eigen, eos.

Библиотека OpenCV [30] – библиотека с открытым исходным кодом для реализации программ, использующих компьютерное зрение. Включает в себя весьма обширный функционал: ввод и вывод изображений и видео, обработка изображений (применение фильтров, геометрические преобразование, работа с цветовыми пространствами), калибровка камеры, поиск соответствующих точек на нескольких изображениях, распознавание образов и многое другое. Разработана на языке C++, сейчас активно переносится и на другие языки программирования.

Библиотека Eigen [31] – библиотека линейной алгебры с открытым исходным кодом. Используется для ускорения и упрощения векторных и матричных вычислений. Написана для языка C++.

Библиотека eos [32] – библиотека для работы с трёхмерной морфируемой моделью с открытым кодом. Поддерживает базовые функции для генерации лиц с помощью морфируемой модели, а также функции калибровки камеры и функции, необходимые для подгонки морфируемой модели. Также присутствует функционал работы с текстурами. Данная библиотека поддерживает работу с тремя морфируемыми моделями, это Surrey Face Model (SFM) [33] – это морфируемая модель, полученная в результате лазерного сканирования 170 различных людей, имеет четыре разрешения – 1k, 3k, 16k и 29k вершин, также, помимо формы, предоставляются и текстуры; 4D Face Model (4DFM) [34] – трёхмерная морфируемая модель, построенная по результатам сканирования нескольких сотен лиц людей разных национальностей и возрастов, помимо лицевых особенностей, данная модель может генерировать различные эмоции, 36 разных выражений лица, модель генерирует не только лицо, а всю голову целиком; Basel Face Model (BFM) [35] – трёхмерная морфируемая модель, полученная сканированием 100 женщин и 100 мужчин, в 2017 году была выпущена уже вторая версия этой базы данных, первая в 2009 году, обе версии поддерживаются библиотекой. Библиотека написана на языке C++, но также часть функционала доступна на языках Python и Matlab.

3.3 Описание алгоритма

В качестве входных данных я буду использовать базу данных 300-W [36][37][38], состоящую из 300 фотографий различных людей и соответствующих этим фотографиям списков 68 ключевых точек лиц, расположенных, как показано на Рис.1. В данной базе представлены фотографии с различными эмоциями, различные положения лиц, лица людей разных национальностей и возрастов. То есть в программу на вход подаётся одна фотография лица и список расположения 68 ключевых точек на фотографии.

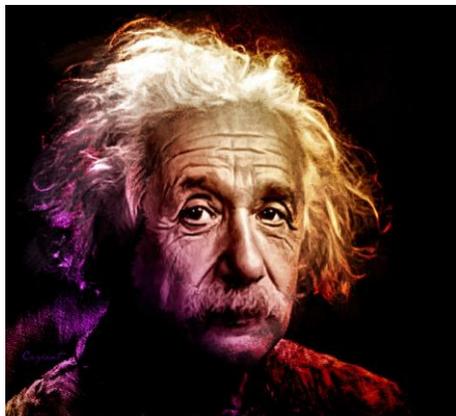
После загрузки входных данных необходимо определить положение камеры относительно лица, изображенного на фотографии. Так как ключевые точки лица уже известны, необходимо соотнести их с трёхмерной морфируемой моделью, тем самым оценив позу, в которой находится лицо на изображении. Для решения этой задачи нормализуем ключевые точки входной фотографии и имеющейся морфируемой модели. Проведём преобразование подобия, заключающееся в переносе центра и масштабировании. Теперь можно вычислить параметры камеры, решив систему линейных уравнений.

Следующий этап заключается в подборе параметров трёхмерной морфируемой модели, подходящих к лицу, изображенному на фотографии. Этот этап подробно описан в разделе 2.2, необходимо минимизировать функцию: $E(\vec{\alpha}_{id}, \vec{\alpha}_{exp}, \vec{\beta}, \vec{p}) = E_{pix} + c_{lands} \cdot E_{lands} + c_{reg} \cdot E_{reg} \rightarrow min.$

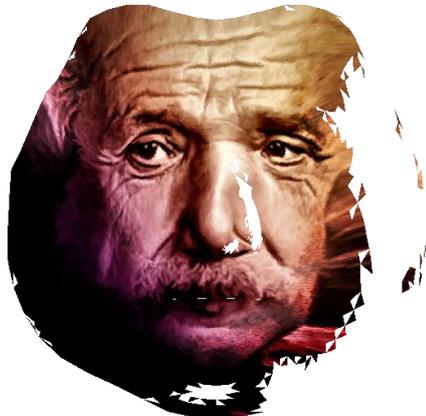
Подгонка трёхмерной морфируемой модели осуществлена, теперь необходимо произвести извлечение текстуры из входного изображения. Так как трёхмерная модель построена, то мы имеем соответствие между точками, описывающими модель и соответствующими им пикселями на фотографии. Поэтому можно легко получить текстуру лица.

На выходе программа даёт три файла: .obj – трёхмерная модель лица,

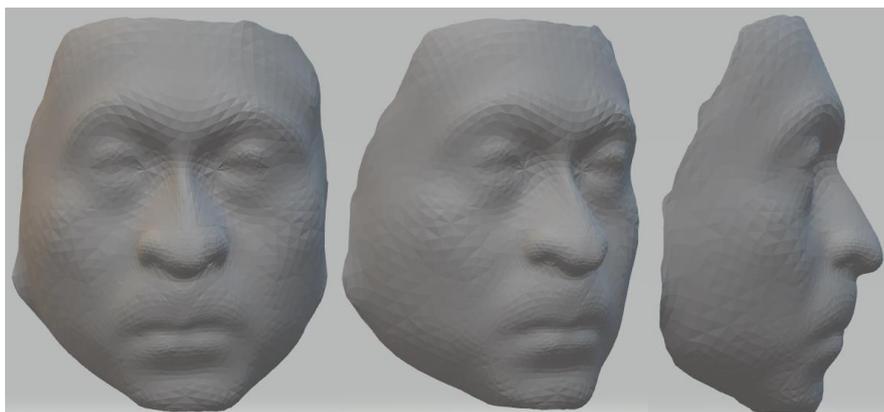
.isomap – извлеченная текстура лица и .mtl – файл, содержащий соответствия текстуры и трёхмерной модели, правила, по которым текстура должна накладываться на 3D-объект. Ниже приведён пример результата работы программы:



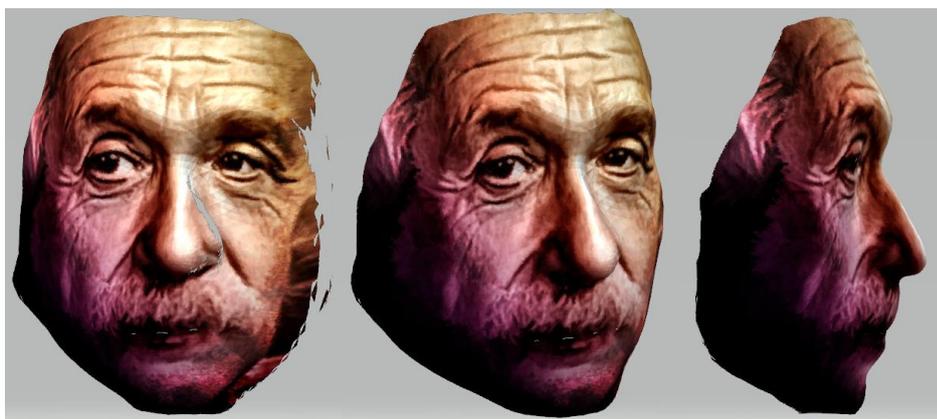
(a)



(б)



(в)



(г)

Рис. 2: (a) Входное изображение; (б) Извлечённая текстура; (в) Трёхмерная модель без текстуры; (г) Трёхмерная модель с текстурой;

3.4 Эксперименты

Проведём тестирование полученного алгоритма на различных лицах из базы данных 300-W, будем выбирать фотографии с различными лицевыми особенностями, чтобы выявить недостатки и преимущества данного алгоритма.

Как видно на Рис.3, наличие на лице очков не мешает процессу построения трёхмерной модели, но и в модели они никак не учитываются, очки входят в текстуру вместе со всем остальным лицом.



Рис. 3: Пример работы программы при наличии на лице дополнительных аксессуаров

Алгоритм достаточно хорошо справляется с лицами, имеющими не нейтральное выражение, выражающими какую-то эмоцию, пример Рис.4.



Рис. 4: Пример работы программы с ярко выраженной эмоцией на лице

Фотографии с нейтральной эмоцией дают также хорошие результаты, пример Рис.5.

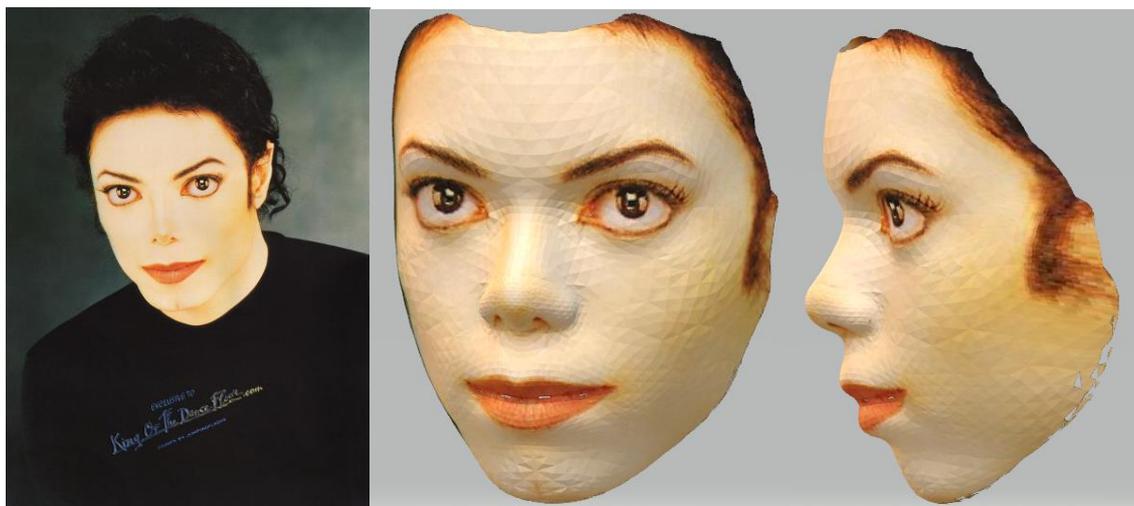


Рис. 5: Пример работы программы с нейтральной эмоцией на лице

Алгоритм успешно работает с лицами людей разных возрастов, например, успешно построена модель детского лица Рис.6.



Рис. 6: Пример работы программы с детским лицом

Программа справляется даже с лицами, имеющими ярко выраженную асимметрию:



Рис. 7: Пример работы программы с лицом с ярко выраженной асимметрией

При сокрытии части лица алгоритм выдаёт следующие результаты, представленные на Рис.8.



Рис. 8: Пример работы программы с фотографией, где часть лица скрыта

В целом можно считать, что алгоритм успешно справляется с поставленной задачей, имеются мелкие недочеты, но они не являются критичными.

ЗАКЛЮЧЕНИЕ

В рамках данной научно-исследовательской работы были изучены общие принципы восстановления трёхмерных сцен по фотографиям, были выделены и изучены характерные черты, присущие восстановлению трёхмерных моделей лица. Были рассмотрены популярные на данный момент современные тенденции и методы решения данной задачи, а именно методы реконструкции лиц на основе только одной фотографии с использованием искусственных нейронных сетей и подходы, использующие трёхмерную морфируемую модель.

Был самостоятельно реализован алгоритм, основанный на подгонке трёхмерной морфируемой модели, выполняющий реконструкцию трёхмерной модели лица с использованием одного изображения. Для реализации алгоритма использовались библиотеки OpenCV, Eigen, eos. Алгоритм был реализован на языке C++. В качестве результата алгоритм выдаёт три файла: .obj – трёхмерная модель лица, .isomap – извлеченная текстура лица и .mtl – файл, содержащий соответствия текстуры и трёхмерной модели, правила, по которым текстура должна накладываться на 3D-объект. Был проведён ряд экспериментов, оценивающий качество полученного алгоритма, выявлены его недостатки.

Для улучшения качества получаемой трёхмерной модели можно попробовать использовать морфируемую модель с большим числом полигонов. Также можно добавить алгоритм заполнения пробелов в текстурах.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Кирпичников, А. П. Трёхмерная реконструкция сцены по нескольким изображениям // А. П. Кирпичников, И. И. Шамсутдинов, М. П. Шлеймович
- [2] Тупицын, И. В. Реконструкция трехмерной модели объекта на основе стереопары при решении задач 3d-моделирования
- [3] Moravec, H. Rover visual obstacle avoidance // In International Joint Conference on Artificial Intelligence, Vancouver, Canada, 1981, pp. 785-790.
- [4] Leibe, B. Robust Object Detection with Interleaved Object Categoization and Segmentation Springer Science + Business Media // B. Leibe, A. Leonardis, B. Schiele, LLC. – 2007.
- [5] Papageorgiou, C. A trainable system for object detection Computer Vision. // C. Papageorgiou, T. Poggio – Vol.38. No.1, 2000. – pp. 15-33
- [6] Lowe, D. G. Distinctive Image Features from ScaleInvariant Keypoints // International Journal of Computer Vision, 60, 2, с. 91-110, 2004.
- [7] Bay, H. SURF: Speeded Up Robust Features // H. Bay, A. Ess, T. Tuytelaars, L. V. Gool – Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, с. 346–359, 2008
- [8] Конушин, А. Геометрические свойства нескольких изображений. // Компьютерная графика и мультимедиа. Выпуск №4(3)/2006.
- [9] Tsai, R. Y. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. – Miami Beach (FL). – 1986. – P. 364-374
- [10] Zhengyou Zhang IEEE Transactions on Pattern Analysis and Machine Intelligence // December 2000, Vol 22: pp. 1330-1334
- [11] Lourakis, M. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm// M. I. A. Lourakis , A. A. Argyros – Rep. / Technical Report 340, Institute of

- Computer Science FORTH, Heraklion, Crete, Greece, 2004.
- [12] Татаренков Д. А. Анализ методов обнаружения лиц на изображении // Молодой ученый. — 2015. — №4. — С. 270-276. — URL <https://moluch.ru/archive/84/15524/> (дата обращения: 29.05.2019).
- [13] Orman A Study On Face, Eye Detection And Gaze Estimation // Orman, Zeynep & Battal, Abdulkadir & Kemer, Erdem – International Journal of Computer Science & Engineering Survey.
- [14] Yang, G. Human face detection in a complex background. Pattern Recognition // G. Yang and Thomas S. Huang, 27(1):53–63, 1994.
- [15] Kotropoulos, C. Acoustics, Speech, and Signal Processing // C. Kotropoulos, I. Pitas, 1997. ICASSP-97, 1997 IEEE International Conference on p.2537–2540 v. 4
- [16] Leung, T. K. Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching // T. K. Leung, M. C. Burl, P. Perona
- [17] Yow, K. C. Feature-based human face detection// K. C. Yow, R Cipolla, Image and vision computing 15 (9), p. 713–735, 1997
- [18] Sinha, P. Perceiving and Recognizing threedimensional forms // PhD thesis, Massachusetts Institute of Technology.
- [19] Lanitis, A. Image Anal. Classifying variable objects using a flexible shape model // C.J. Taylor, T. Ahmed, T.F. Cootes, Wolfson, Image Processing and its Applications, 1995., p.70–74
- [20] Мак-Каллок, У. С. Логическое исчисление идей, относящихся к нервной активности // У. С. Мак-Каллок, В. Питтс, Автоматы / Под ред. К. Э. Шеннона и Дж. Маккарти. — М.: Изд-во иностр. лит., 1956. — С. 363—384. (Перевод английской статьи 1943 г.)
- [21] Jackson, A. S. Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression. // A. S. Jackson, A. Bulat, V. Argyriou, G. Tzimiropoulos. IEEE International Conference on Computer Vision (ICCV), 2017
- [22] Jackson, A. S. Backpropagation Applied to Handwritten Zip Code

- Recognition // A. S. Jackson, A. Bulat, V. Argyriou, G. Tzimiropoulos., Neural Computation, 1(4):541-551, Winter 1989.
- [23] Matusugu Subject independent facial expression recognition with robust face detection using a convolutional neural network // Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda Neural Networks. 16 (5): 555—559.
- [24] Zhu, X. Face alignment across large poses: A 3d solution. // X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li., arXiv:1511.07212, 2015
- [25] Tran, L. On Learning 3D Face Morphable Model from In-the-wild Images. // L. Tran, X. Liu, arXiv:1804.03786, 2018
- [26] Blanz, V. A morphable model for the synthesis of 3D faces. // V. Blanz, T. Vetter., SIGGRAPH '99 Proceedings of the 26th annual conference on Computer graphics and interactive techniques, 1999. P. 187–194.
- [27] Romdhani, S. Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. // S. Romdhani, T. Vetter., 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. Vol. 2, P. 986 – 993.
- [28] Kemelmacher-Shlizerman, I. 3d face reconstruction from a single image using a single reference face shape. // I. Kemelmacher-Shlizerman, R. Basri, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011. Vol. 33, No 2. P. 394–405.
- [29] Microsoft Visual Studio. Official website. — URL <https://visualstudio.microsoft.com/ru/>
- [30] Библиотека OpenCV с открытым исходным кодом — URL <https://opencv.org/>
- [31] Библиотека Eigen с открытым исходным кодом — URL http://eigen.tuxfamily.org/index.php?title=Main_Page
- [32] Библиотека eos с открытым исходным кодом — URL <https://github.com/patrikhuber/eos>
- [33] Surrey Face Model — URL

<https://www.cvssp.org/faceweb/3dmm/facemodel/>

- [34] 4D Face Model — URL <https://www.4dface.io/4dfm/>
- [35] The Basel Face Model — URL <https://faces.dmi.unibas.ch/bfm/>
- [36] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. 300 faces In-the-wild challenge: Database and results. Image and Vision Computing (IMAVIS), Special Issue on Facial Landmark Localisation "In-The-Wild". 2016.
- [37] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge. Proceedings of IEEE Int'l Conf. on Computer Vision (ICCV-W), 300 Faces in-the-Wild Challenge (300-W). Sydney, Australia, December 2013.
- [38] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. A semi-automatic methodology for facial landmark annotation. Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR-W), 5th Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2013). Oregon, USA, June 2013.