Дипломная работа

# Введение

Задача идентификации по биометрическим признакам сегодня имеет большое значение во многих сферах применения. Особенно, важным является при разграничении доступа к важным хранилищам информации или субъектам, которые должны быть доступны лишь ограниченному кругу лиц.

Целью данной работы является создание системы идентификации по 3D изображиню лица человека.

Для реализации системы были поставлены следующие задачи:

1. Сбор данных.
2. Обработка полученных данных
3. Дообучение нейронной сети на данных
4. Обработка результата

# Теоретическая часть

## Наборы данных.

В каччестве набора данных для создания нейронной сети используются наборы следующих данных:

1. 3DTEC
2. BU\_3DFE
3. Bosphorus.

3DTEC (<http://i14s50.anthropomatik.kit.edu/download/BEFIT_Notre_Dame_3DTEC_v2.pdf>)

В базе данных 3DTEC Имеется 440 изображения в формате ABS.

Чтобы иметь возможность работать с изображениями, сначала необходимо изменить формат каждого изображения на PLY. Что было получено с использованием языка программирования Python и библиотеки NuPy.

Данная база данных была разработана для использования в задачах по распознаванию близнецов.

***BU-3DFE*** (<http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html>)

В данной базе располагается 2500 изображений 3D лиц более чем 100 человек, которые были созданы для задач распознавания и анимации.

Соотношение изображений мужчин и женщин составляет 56% к 44%. Возраст актёров от 18 до 70 лет. Содержит множество рас:

* Европейцы
* Африканцы
* Среднеазиатская
* Индийцы
* Испанолатинская. (Hispanic Latino)

Каждый субъект исполнил 7 различных выражений лица перед 3D-сканером

Изображены следующие виды эмоций:

* Счастье
* Отвращение
* Страх
* Злость
* Удивление
* Грусть
* Нейтральная

База данных BOSPHORUS специально создана чтобы проводить исследования в области 3D и 2D обработки изображений лиц людей, а также 3D реконструкции лиц. База состоит из изображений 105 человек, и всего 4666 изображений 3D облаков. В базе содержатся не только нормализованные изображения, но также:

1. Выражение эмоций – до 35 для отдельного человека.
2. Подсчитанные лицевые движения.
3. Третья часть базы создана с привлечением профессиональных актёров.

Различные уровни поворота лица.

# FaceWarehouse (<http://kunzhou.net/zjugaps/facewarehouse/>)

Cao Chen, Yanlin Weng, Shun Zhou, Yiying Tong, Kun Zhou: "FaceWarehouse: a 3D Facial Expression Database for Visual Computing", IEEE Transactions on Visualization and Computer Graphics, 20(3): 413-425, 2014

База данных 3D изображений лиц с различными выражениями для приложений. Содержит изображения 150 человек в возрасте от 7 до 80 лет разных этнических групп. Для получения данных использовалась RGBD камера Kinect.

Данные для каждого человека хранятся в отдельной директории. В каждой директории "Tester\_xx",хранятся RGBD данные, отмечены 74 особые точки, the reconstructed meshes, and user-specific blendshapes for this person.

* Tester\_1: folder name with person 1
  + TrainingPose
    - pose\_0.pose - pose\_20.pose: данные от Kinect RGBD камеры
    - pose\_0.obj - pose\_20.obj:реконструированные меши.
    - pose\_0.land - pose\_20.land: отмеченные 2D точки для первого кадра в директории.
  + Blendshape
    - shape.bs: двоичный файл с выражением blendshape
    - shape\_0.obj - shape\_46.obj: выражения blendshapes

Примеры возможных применений:

* перенос эмоций.
* создание анимаций изображений в реальном времени.
* создание новых эмоций для отдельного персонажа используя совмещение имеющихся.

Bosphorus Database

База данных BOSPHORUS специально создана чтобы проводить исследования в области 3D и 2D обработки изображений лиц людей, а также 3D реконструкции лиц. База состоит из изображений 105 человек, и всего 4666 изображений 3D облаков. В базе содержатся не только нормализованные изображения, но также:

1. Выражение эмоций – до 35 для отдельного человека.
2. Подсчитанные лицевые движения.
3. Третья часть базы создана с привлечением профессиональных актёров.

Различные уровни поворота лица.

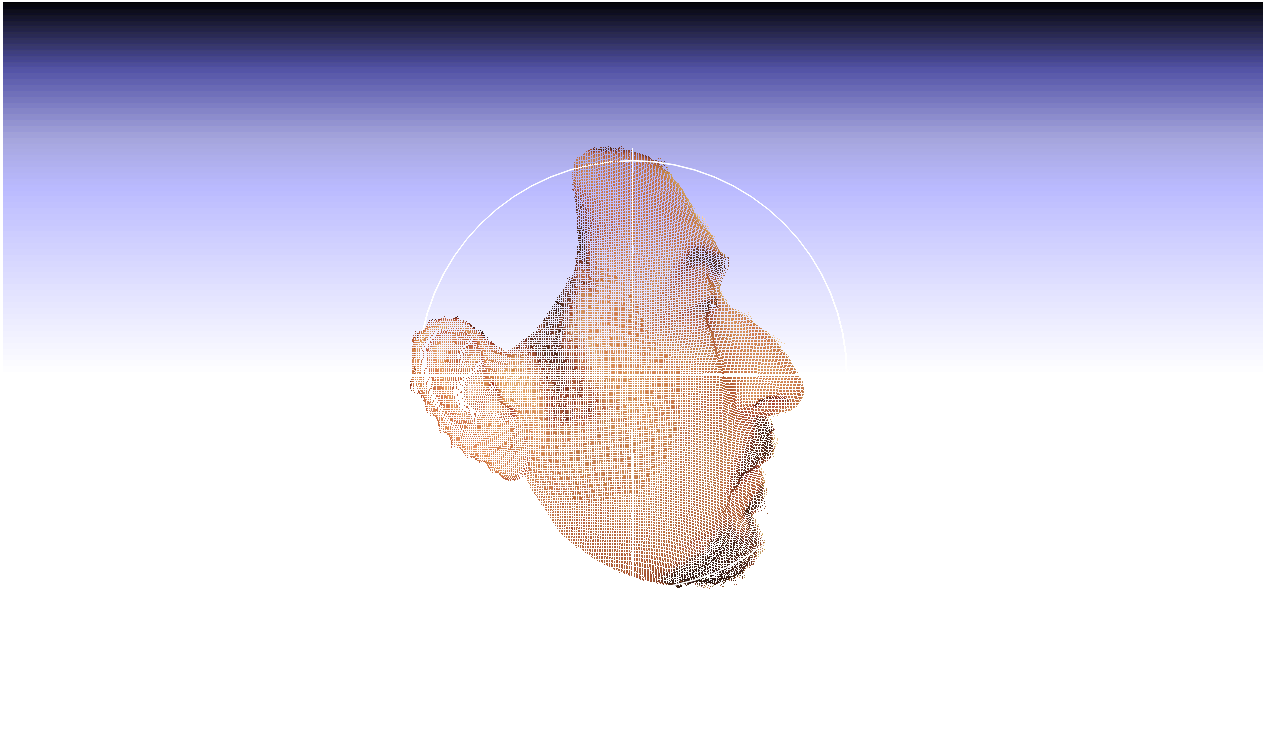


Рис 1. Пример изображения bs002\_YR\_L90\_0.ply до процедуры поворота.

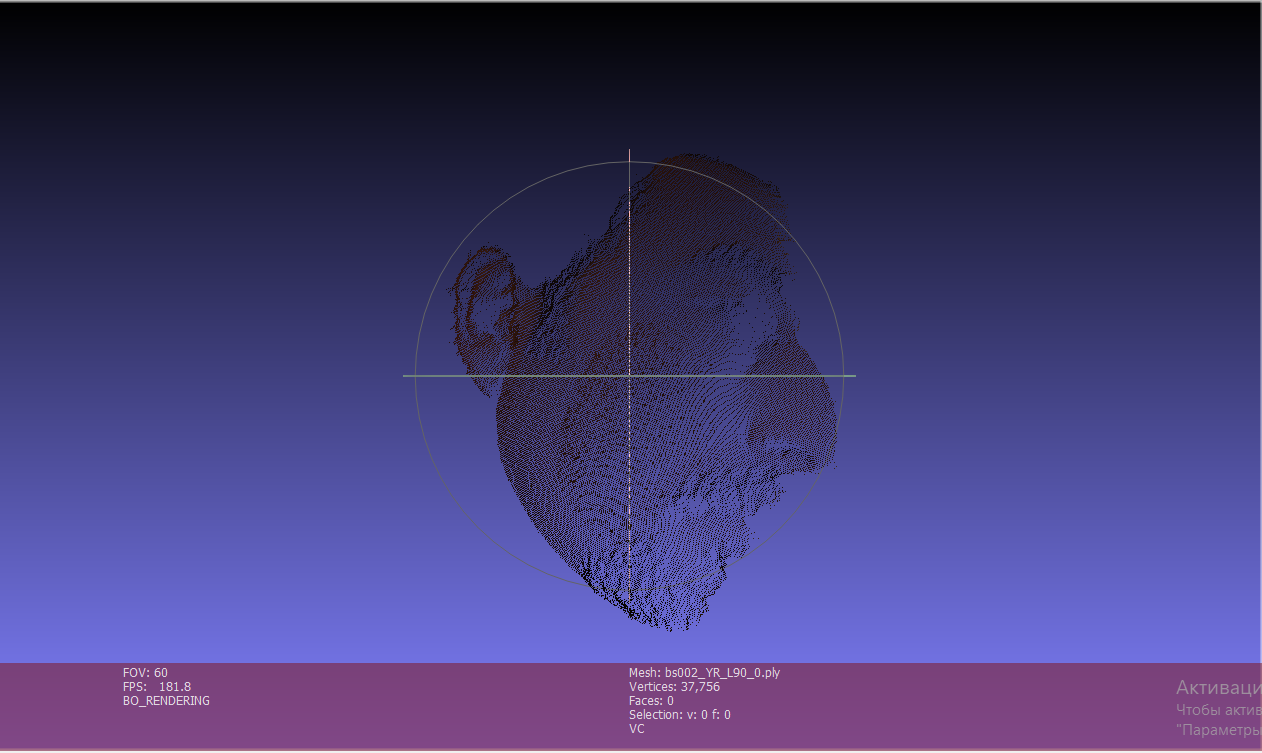


Рис 2. Пример изображения bs002\_YR\_L90\_0.ply после процедуры поворота.

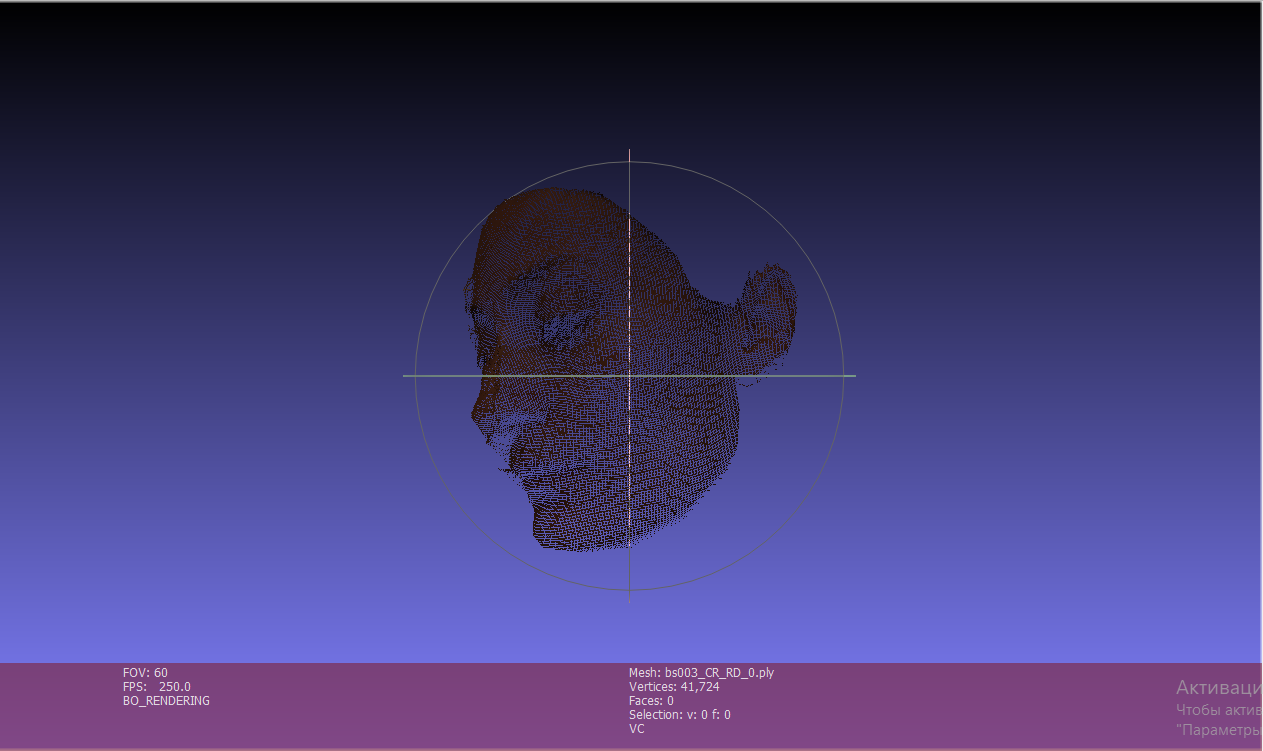


Рис 3. Пример изображения bs003\_CR\_RD\_0.ply после процедуры поворота.

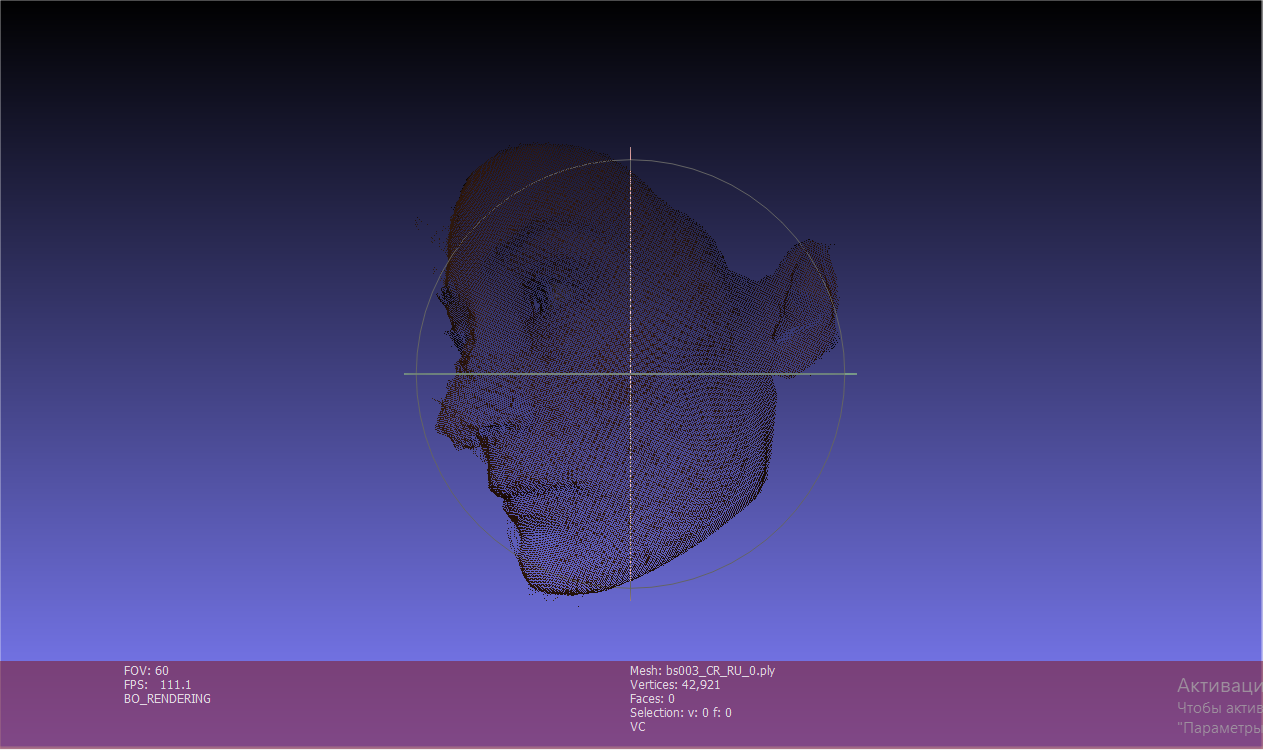


Рис 4. Пример изображения bs003\_CR\_RD\_0.ply после процедуры поворота.

VGG Face 2

В качестве исходного набора данных для протестированных архитектур нейронных сетей используется VGG Face 2 (<http://www.robots.ox.ac.uk/~vgg/data/vgg_face2/index.html>).

<http://www.robots.ox.ac.uk/~vgg/publications/2018/Cao18/cao18.pdf>

VGG Face 2 собран научной группой Оксфордского Университета (*University of Oxford). Набор содержит более 3.3 миллионов изображений лиц более чем 9000 субъектов. Размеры изображений варьируются. Всем изображенями соответствуют различные уровни аугментации, такие как: различные положения лица, присутствие посторонних лиц, использование дополнительных средств сокрытия элементов лица (очки, шляпы и т.д.) Набор был*

** **

*Рис. 3 Примеры изображения лица субъекта в наборе данных VGG Face 2.*

** ** 

*Рис. 3 Примеры изображения лица субъекта в наборе данных VGG Face 2.*

## Задача идентификации по 3D изображению лица.

Назовём режимом идентификации следующий эксперимент. Создадим два набора данных – Gallery и Probe.

Стандартным способом проведения идентификации является расположения в Gallery нормализованных изображений, а в Probe произвольного набора изображений.

На выходе нейронной сети получается набор признаков. Набор признаков представляется в виде вектора, длина которого определяется конкретной архитектурой нейронной сети. Например:

* для сети VGG16 набор признаков слоя fc7 имеет длину 4096;
* для сети ResNet-50 вектор признаков слоя classifier имеет длину 8631.

Для того, чтобы сравнивать похожесть изображений в задаче идентификации мы используем косинусное расстояние, вычисляемое для двух векторов признаков, получаемых на выходе нейронной сети.

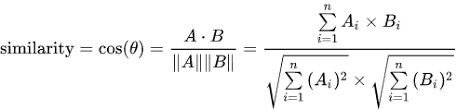


Рис 1. Косинусное расстояние между векторами A и B.

Подобный подход позволяет освободиться от создания классификатора для фиксированного набора изображений и вычислять сходство между произвольным количеством субъектов идентификации.

В нормализованной группе находились изображения нормальные (\*\_N\_N\_0.ply) изображения для каждого из объектов идентификации. Затем в 105 экспериментах для каждого отдельного объекта выбирались оставшиеся, кроме нормального, изображения и проводилось их сравнение нейронной сетью. На выходе нейронной сети мы получаем вероятности совпадения изображения из нормализованной и тестовой групп.

*.*

## Нейронные сети.

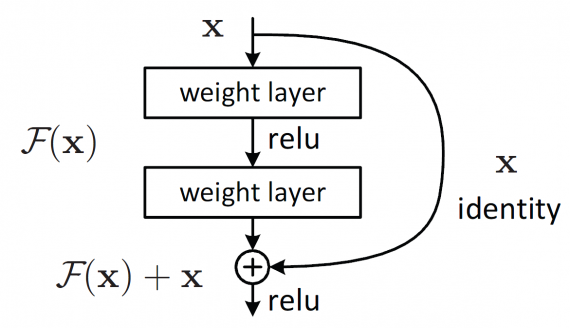
*Архитектура Resnet-50.*

*В предыдущих работах была исследована нейронная сеть с VGG16 обученнная на наборе данных VGGFace. (*<https://www.robots.ox.ac.uk/~vgg/data/vgg_face/>)

*Данная нейронняа сеть является сверхглубокой. Имеет 22 слоя, включая полносвязные и свёрточные слои. На вход сети подаётся изображение размерности 224\*224 пикселей*

*Архитектура ResNet также является сверхглубокой нейронной сетью. Различные модификации могут включать различные наборы слоёв. Отличием ResNet от других свёрточных нейронных сетей является наличие специальных Residual слоёв.*

*Схематиченое изображение Residual слоя изображено на рисунке Рис. 2. В задачах классификации изображений архитектуры ResNet неоднократно доказывали свою эффективность в различные времена являясь state-of-art решениями.*



*Рис. 2. Схематичное изображение Residual блока. Такой блок состоит из суммы вектора данных X и вектора обработанных данных преобразованием F(X).*

*Использование ResNet-50 было определено тем, что существуют несколько свободно распространяемых предобученных на VGG Face 2 нейронных сетей данной архитектуры.*

*Описание предобученных архитектур нейронных сетей ResNet-50.*

*(*<https://github.com/ox-vgg/vgg_face2>)

| **Архитектура** | **Размер выходного слоя** | **TAR@FAR = 0.001** | **TAR@FAR = 0.01** | **Фреймворк** |
| --- | --- | --- | --- | --- |
| ResNet-50 | 8631 | 0.891 | 0.947 | Caffe, PyTorch |
| ResNet-50-256D | 256 | 0.898 | 0.956 | Caffe, PyTorch |
| ResNet-50-128D | 128 | 0.904 | 0.956 | Caffe, PyTorch |

*Каждая из представленных выше моделей имеет 228 слоёв.*

*Различия заключаются лишь в количестве нейронов выходного слоя.*

# Практическая часть

## Реализация Transfer Learning.

Процесс распознавания.

Транслирование наборов данных из различных форматов в облако точек.

В связи с тем, что программные системы применяют разные форматы хранения 3D изображений, вначале их следует приводить в один формат данных.

Для этих целей разработан набор скриптов. На последнем шаге приводящие данные в формат PLY. Применяется язык программирования Python версии 3.7 с библиотеками для научных вычислений (NumPy, SciKit-learn) и реконструирования 3D-изображений (PCL).

Шаг 1. Применение набора скриптов для преобразования данных в формате PLY из других форматов, таких как (OBJ, WRL, ABS).

Шаг 2. Преобразование PLY облаков точек в двоичный формат NPY, содержащий представление 2D проекции облака точек на плоскость XY.

Для этого используются файлы Preprocessing.py и XYProjection.py

На вход нейронной сети передаются данные в двоичном формате, получаемые из файлов облаков точек PLY.

Конфигурация тестовой системы: CPU = intel i3 9100F, GPU = NVIDIA RTX 2060, RAM = 16 GB.