

**Федеральное государственное автономное образовательное
учреждение высшего образования «Национальный
исследовательский университет «Высшая школа экономики»**

Факультет компьютерных наук
Основная образовательная программа
"Прикладная математика и информатика"

**ОТЧЕТ ОБ ИССЛЕДОВАТЕЛЬСКОМ ПРОЕКТЕ
НА ТЕМУ "ВИЗУАЛИЗАЦИЯ МНОГОМЕРНЫХ ДАННЫХ С
ИСПОЛЬЗОВАНИЕМ ЛИЦ ЧЕРНОВА"
(ПРОМЕЖУТОЧНЫЙ, ЭТАП 1)**

Выполнил студент:
Веселов Илья Николаевич, группа 218, 2 курс

Принял руководитель проекта:
Заведующий Базовой кафедрой Тинькофф Департамента больших дан-
ных и информационного поиска, Доцент,
Иванов Андрей Александрович

Москва 2023

Содержание

1	Основные термины, определения и сокращения	3
2	Введение	4
2.1	Цель и задачи	5
2.2	Обзор источников	6
3	Разработка и реализация	12
3.1	Геометрическая модель	12
3.2	Реализация в GeoGebra	14
3.3	Визуализация в GeoGebra	18
3.4	Реализация модели в Python	20
3.4.1	Основные функции	21
4	Визуализация датасета Iris	23
4.1	Описание датасета Iris	23
4.2	Использование библиотеки для визуализации	24
4.3	Анализ Результатов	25
5	Заключение	26
5.1	Результаты	26
5.2	Будущее развитие	27
6	Календарный план работы над проектом	29

1 Основные термины, определения и сокращения

Визуализация - это общее название приёмов представления данных (информации) или физических явлений в виде, который удобен для зрительного наблюдения и последующего анализа. [12]

Визуализация данных — это графическое представление информации и аналитики: графики, диаграммы, карты, дашборды. [13]

Дашборд - это интерактивная аналитическая панель, графический интерфейс. [14]

Пиктографика - это технология визуализации информации, в которой используются графические изображения для представления и улучшения восприятия многомерных данных. [15]

Лица Чернова - это способ отображения многомерных данных, придуманный прикладным математиком, статистиком и физиком Германом Черновым в 1973 году, в виде человеческого лица и его отдельных частей, таких как глаза, уши, рот и нос, которые представляют значения переменных по их форме, размеру, расположению и ориентации. [5]

2 Введение

Задача визуализации данных насчитывает десятки тысяч лет. Это и наскальные изображения, и современная картография, и навигация в робототехнике, и исследования томографии, экономики отдельных стран и регионов, и системы поиска, хранения и анализа информации в компьютерной сети, и распознавание образов - вот диапазон задач, которые объединены общим понятием визуализации. Использование графического отображения данных в исследовательских трудах значительно увеличивает скорость передачи и уровень восприятия сложной информации. Это также помогает развивать интуицию и образное мышление, которые необходимы для всех специалистов.

Лучше всего у человека сформировано умение восприятия лиц других людей. Информация с эмоциями является основным критерием для принятия решений и реализации различных действий. Тем не менее человек нуждается не только в объективных данных, но и в получении субъективной информации, которую можно получить либо во время общения с человеком, либо от технической системы, которая будет синтезировать эмоции, либо через «лица Чернова». Пример использования лиц Чернова изображен на рисунке [2.1](#).

Объектом исследования в данной работе является метод визуализации многомерных данных "лица Чернова" в котором точки многомерного пространства изображаются в виде схематических человеческих лиц, которые конструируются по определенным правилам, так что каждый из конструктивных параметров лица соответствует одному параметру изучаемого объекта.

Однако относительно малое количество программных пакетов и средств предлагают возможность построения и интерпретации пиктограмм Чернова.

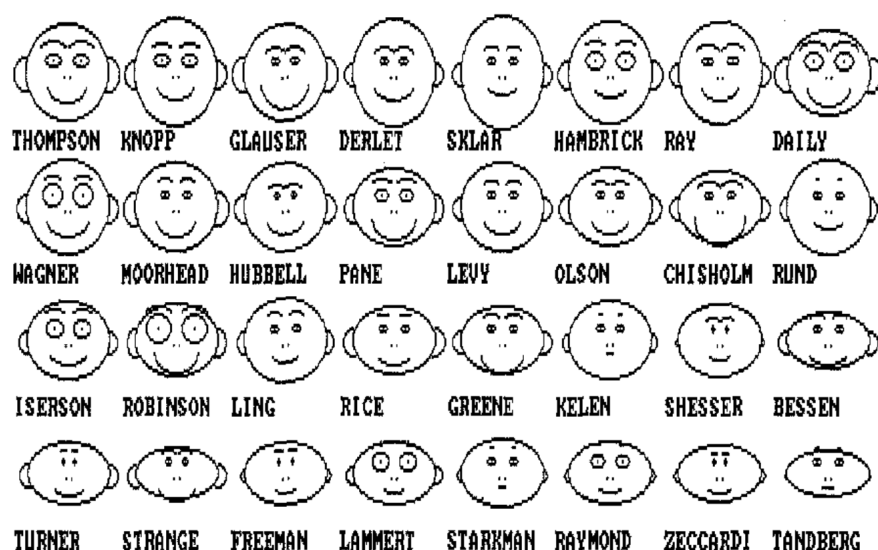


Рис. 2.1: На изображении продемонстрированы лица Чернова, нарисованные схематичные лица, которые отображают оценки 32 претендентов в ординатуру по неотложной медицинской помощи. Кандидаты, которые, по прогнозам, будут самыми сильными, расположены вверху слева, а самые слабые - внизу справа. Высота головы соответствует общей оценке, размер глаз - оценке комиссии, размер рта - оценке за собеседование, размер ушей - оценке за письмо, а брови - оценке за автобиографию. Для каждой переменной более крупные объекты дают лучшие результаты. [4]

2.1 Цель и задачи

Ввиду вышеизложенного цель данного исследовательского проекта следующая: исследование существующих методов визуализации, сравнение их с методом генерации лиц Чернова, разработка и реализация новых программных средств на языке программирования Python.

Для достижения цели поставлены следующие задачи:

- изучение принципов работы метода визуализации данных с использованием лиц Чернова,
- реализация MVP библиотеки визуализации данных,
- сравнение с другими методами визуализации многомерных данных,
- анализ возможностей, ограничений и недостатков визуализации средствами лиц Чернова,

- исследование влияния размера выборки и количества измерений на точность и эффективность данного метода,
- разработка заключения и выдвижения перспектив для дальнейшего развития и изучения метода визуализации лица Чернова.

2.2 Обзор источников

Важное место в области исследования метода визуализации данных с применением лиц Чернова занимает труд [5] “The use of faces to represent points in k-dimensional space graphically”, автор Герман Чернов. Из этой статьи становится известен новый метод представления многомерных данных. В данном методе применяется 18 параметров лица, а именно: расположение зрачков, изгиб бровей, размер глаз, размер зрачков, наклон глаз, вертикальное расположение глаз, горизонтальное расположение глаз, плотность бровей, горизонтальное расположение бровей, вертикальное расположение бровей, верхняя граница волос, нижняя граница волос, очертание лица, цвет (темнота) волос, наклон штриховки волос, размер носа, размер рта, изгиб рта. Если использовать асимметрию лица, то можно увеличить количество переменных в два раза и сделать равным 36. Пример использования лиц Чернова для визуализации данных показан на Рисунке 2.2.

В течение последних 30 лет специалисты проводили различные исследования этого метода, пытаясь улучшить его возможности для графического представления информации. Две из наиболее известных работ были выполнены Бернардом Флури и Гансом Ридвилем в 1981 году [7] “Graphical Representation of Multivariate Data by Means of Asymmetrical Faces”, в которой была улучшена концепция и добавлена лицам Чернова асимметрия, что увеличило в 2 раза число переменных (до 36), а также Кабуловым Бахрамджаном Тахировичем в 1992 году [8] “A method for constructing Chernoff faces oriented toward interval estimates of the parameters”, в которой автор предло-

1A. FACES FOR 87 FOSSIL SPECIMENS OF EXAMPLE 1

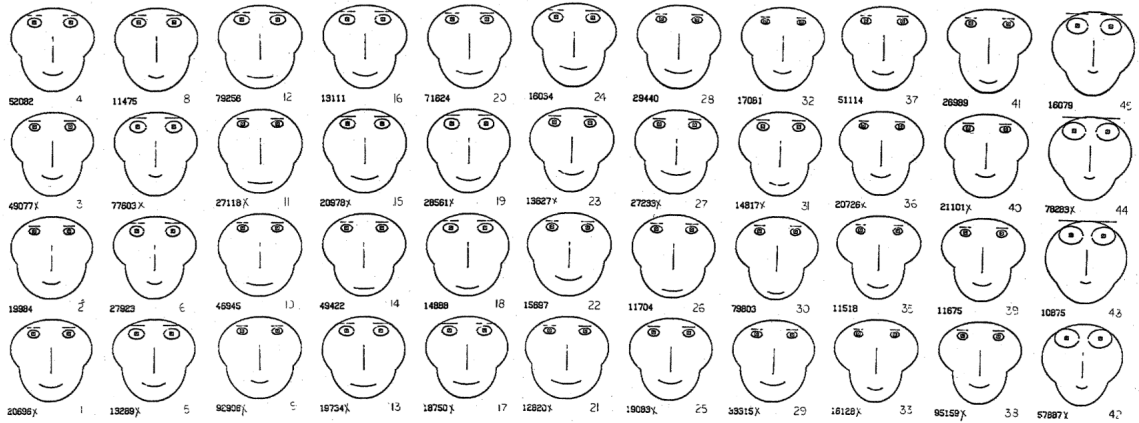


Рис. 2.2: На изображении продемонстрированы лица Чернова, которые созданы на основе данных, собранных путём измерения горных пород. В статье объясняется, что было проведено 8 измерений на каждом из 88 оцифрованных образцов из эоценовой формации жёлтого известняка на северо-западе Ямайки. [5]

жил усовершенствованные правила построения лиц Чернова, позволяющие уменьшать время и увеличивать точность решения задачи оценки состояния данных, а также смог увеличить количество объектов, используемых для отображения многомерных данных.

Всего через четыре года после публикации статьи, написанной Германом Черновым, данный метод визуализации стали внедрять в картографию, используя человеческие лица для представления данных на карте в соответствии с традиционными методами тематического представления. Первым и более известным примером является карта под названием “Жизнь в Лос-Анджелесе, 1970”, разработанная Юджином Тернером (Рисунок 2.3) и составленная Ричардом Доссом из географического факультета Калифорнийского университета в 1977 году. Юджин Тернер [16] написал об этой карте: “Это, вероятно, одна из самых интересных карт, которые я создал, потому что выражение лица вызывает эмоциональную связь с данными”.

Также, в последнее время к идеям визуализации многомерных данных проявили интерес в медицине, где метод визуализации с помощью лиц Чер-

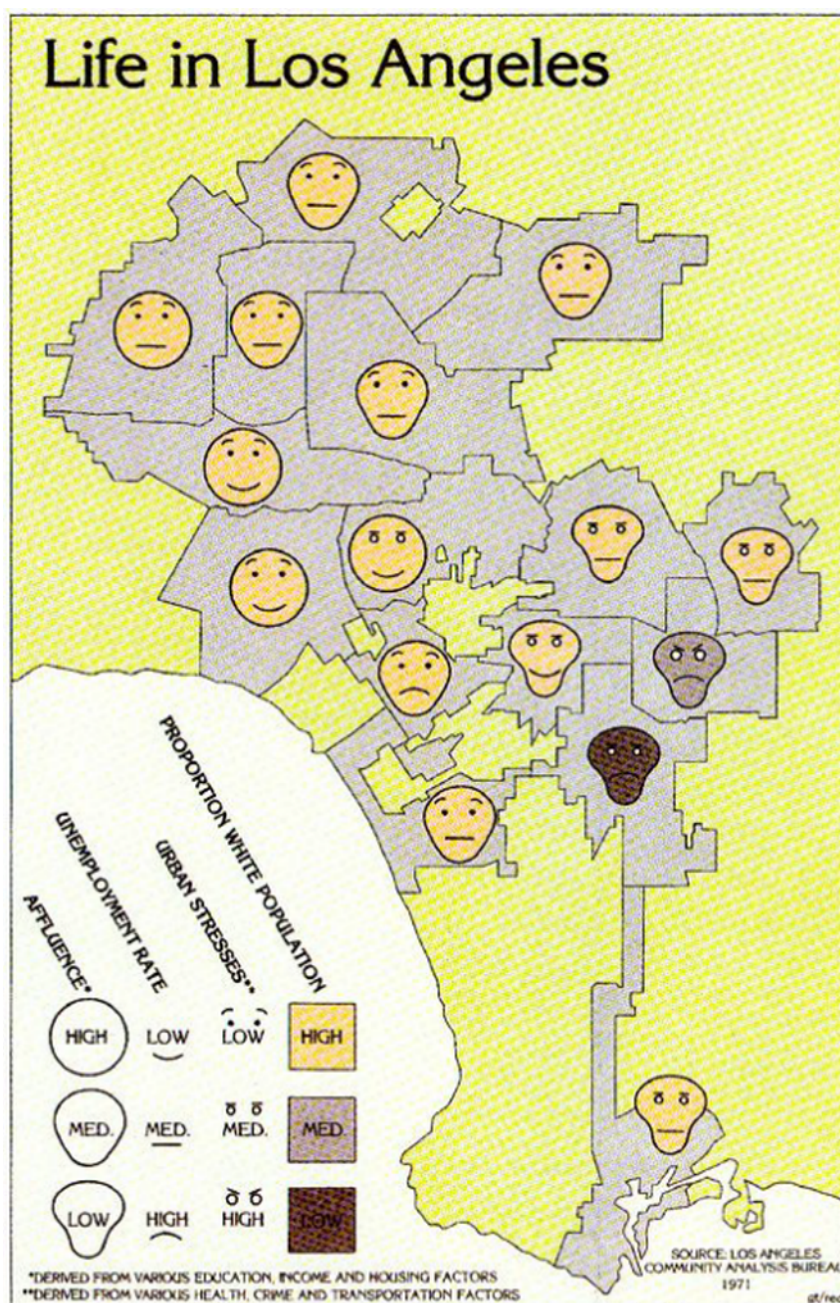


Рис. 2.3: На данной карте, созданной Юджином Тернером, 4 переменные для описания жизни в Лос-Анджелесе 1970 года включают: богатство (зависит от уровня образования, дохода и жилья), уровень безработицы, городской стресс (зависит от уровня здоровья, преступности и развития транспортной системы) и процент белого населения. Эти переменные представлены элементами лица, а именно форма лица, кривизна рта, наклон бровей и цвет лица, соответственно. Созданная карта точно и эстетически отражает условия жизни в районе Лос-Анджелеса. [10]

нова используется для визуализации наблюдений за пациентами. Так, в работе [9] “Анализ многомерных медицинских данных с помощью пиктографов “Лица Чернова””, авторы Осадчая И.А., Берестнева О.Г. и Неме-

ров Е.В. были рассмотрены возможности применения лица Чернова для отображения и анализа медицинских данных на примере задачи выявления изменений физиологических параметров в ответ на психофизиологическое воздействие в зависимости от психологических особенностей больных бронхиальной астмой. В качестве входных данных использовали результаты исследования 2 групп: первую составляли 23 человека, у которых болезнь была связана со стрессом из повседневной жизни, а вторую 29 испытуемых, у которых не было замечено связи обострения бронхиальной астмы со стрессом. На Рисунке 2.4 представлены полученные пиктографики групп пациентов до и после сеанса аудиовизуальной стимуляции, и для построения лиц Чернова использовалось 20 показателей, такие как минутный объем дыхания (овал лица), жизненная ёмкость легких (размер ушей по диагонали), пиковая объемная скорость выдоха (расположение центра рта), общая работа дыхания при спонтанном дыхании (глаз) и так далее. В статье ученые приходят к выводу, что такой способ графического представления позволяет выявить скрытые картины взаимосвязей между данными, которые не могут быть обнаружены с помощью других методов визуализации, а также [9] отмечают, что “понятие “больше – меньше” не всегда соответствует понятию “улучшение – ухудшение” относительно функции организма, что необходимо учитывать при использовании пиктографиков “Лица Чернова” в медицинских научных исследованиях”.

Анализируя вышесказанное, понимаем: визуализация медицинских данных (физические и физиологические данные пациентов) с помощью лиц Чернова эффективна и даёт возможность врачу своевременно определить и оценить состояние пациента в данный момент или результат его лечения в последствии.

Выбор пиктограмм лиц Чернова в качестве визуализации и интерпретации многомерных данных имеет преимущество в том, что человек может легко распознать мимику и её изменения. Однако данный метод визуализа-

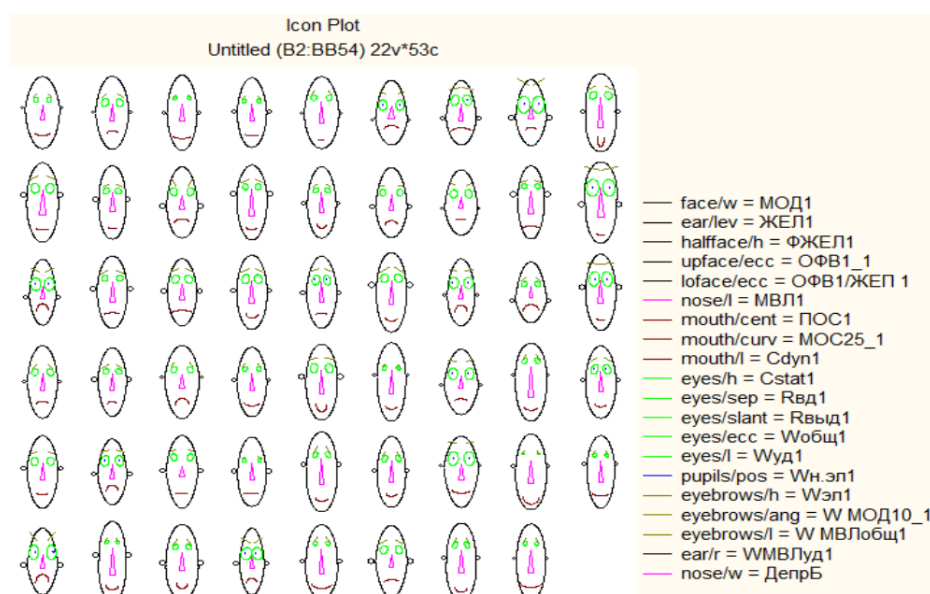


Рис. 1. Отображение состояния больных БА 1 и БА 2 до сеанса ABC

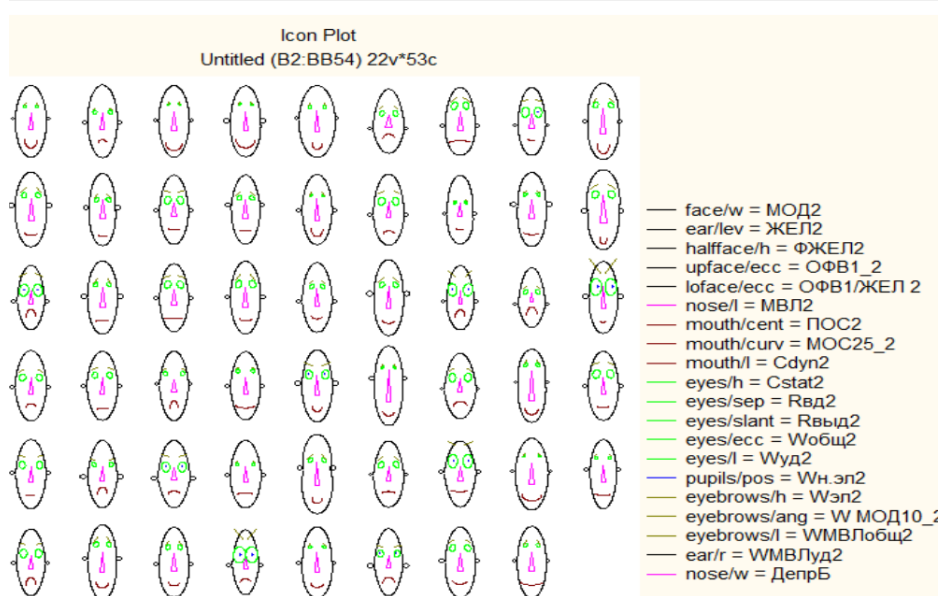


Рис. 2. Отображение состояния больных БА 1 и БА 2 после сеанса ABC

Рис. 2.4: На данном изображении с помощью лиц Чернова продемонстрированы состояния больных бронхиальной астмой из первой и второй группы до и после применения аудиовизуальной стимуляции головного мозга по программе релаксации. [9]

ции имеет свои недостатки, которые заключаются в том, что специалисты могут по-разному интерпретировать одни и те же выражения лица. Также, важен выбор основных параметров изучаемого объекта, потому что неверно выбранные параметры приведут к тому, что исследование объекта станет нерациональным и времязатратным. Кроме того, важно наличие легенды при визуализации данных, иначе будет довольно сложно ин-

терпретировать полученные результаты. Однако, если все эти недостатки учитывать и корректно определять параметры, то эффективность работы с многомерными данными увеличивается по сравнению с применением других существующих моделей.

3 Разработка и реализация

3.1 Геометрическая модель

Лицо Чернова, предложенное Германом Черновым в 1973 году, базируется на 18 переменных x_1, x_2, \dots, x_{18} в соответствующих диапазонах $[0, 1]$. Различные черты лица описываются следующим образом [5]:

- x_1 - радиус до угла лица, $|OP|$
- x_2 - угол OP к горизонту
- x_3 - вертикальный размер лица, $|OU|$
- x_4 - эксцентриситет верхней части лица
- x_5 - эксцентриситет нижней части лица
- x_6 - длина носа
- x_7 - вертикальное положение рта
- x_8 - кривизна рта
- x_9 - ширина рта
- x_{10} - вертикальное положение глаз
- x_{11} - разделение глаз
- x_{12} - наклон глаз
- x_{13} - эксцентриситет глаз
- x_{14} - размер глаз
- x_{15} - положение зрачков
- x_{16} - вертикальное положение бровей
- x_{17} - наклон бровей
- x_{18} - размер бровей

Контур лица представляет собой два эллипса. Из центральной точки O проводится луч к угловой точке P . На противоположной стороне выбирается вторая угловая точка P' так, чтобы OP и OP' имели одинаковую длину и были вертикально симметричны. Верхняя и нижняя части лица представлены соответственно точками U и L , выбираются так, чтобы OU и OL были вертикальными и имели равную длину. Верхняя часть лица

представляет собой эллипс, который определяется точками U, P, P' и x_4 . Нижняя часть определяется аналогично с использованием точек L, P, P' и x_5 . Нос представляет собой вертикальную линию через O , тогда как рот — это дуга окружности. Глаза представлены в виде эллипсов с зрачками, которые расположены на одинаковом горизонтальном расстоянии от центра глаз. Брови — это наклонные отрезки линии над глазами. Оба глаза и брови вертикально симметричны относительно вертикальной линии через O . Для уменьшения влияния x_1 и x_3 , ширина и высота лица нормализуются к двум дюймам. Все параметры, кроме трех эксцентриситетов и кривизны рта, затем нормализуются линейно в диапазоне от 0 до 1. Пример модели с указанными угловыми точками можно увидеть на рисунке 3.1.

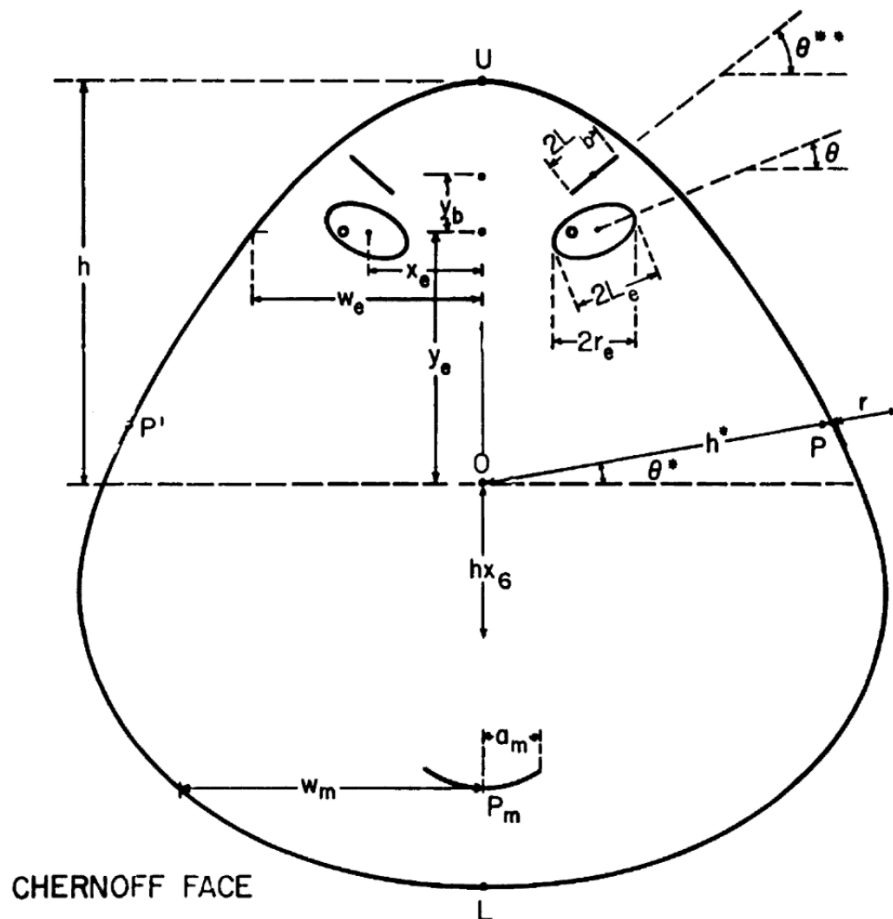


Рис. 3.1: На изображении продемонстрирован пример того, как выполняется построение лица Чернова. Все длины, обозначенные буквами, определяются значениями наблюдений. При этом значения должны пройти предварительную нормировку. [17]

3.2 Реализация в GeoGebra

Переменные модели, такие как x_1, x_2, \dots, x_{18} , требуют специфических преобразований и масштабирований, чтобы подготовить их для использования в GeoGebra. Переменные $x_1, x_2, x_7, x_9, x_{10}$ были созданы с диапазоном значений $[0, 1]$,

- $x_3 \in [-0.95, 0.95]$, • $x_{11} \in [0, 0.2]$, • $x_{16} \in [0, 2]$,
- $x_4 \in [0.25, 1.25]$, • $x_{12} \in [-1, 1]$, • $x_{17} \in [-1, 1]$,
- $x_5 \in [0.2, 2.0]$, • $x_{13} \in [0.05, 1.05]$, • $x_{18} \in [0, 1]$
- $x_6 \in [0.003, 0.303]$, • $x_{14} \in [0.1, 1.1]$,
- $x_8 \in [0.005, 5.005]$, • $x_{15} \in [-0.25, 0.25]$,

Диапазоны значений были скорректированы таким образом, чтобы лица выглядели более «человечными» и чтобы все черты были заметны. Размер глаз был установлен таким образом, чтобы были видны зрачки; длина рта была установлена так, чтобы был виден изгиб. Важно, чтобы все черты были заметны; наличие у лиц человеческих черт — это вопрос предпочтения и соответствия. Возможно, использование человекоподобных лиц поможет в лучшей интерпретации одного набора данных, но в другом будет значительно хуже.

Верхняя и нижняя часть лица Чернова. Эллипсы для верхней и нижней части лица определяются с использованием переменных x_1, x_2, x_3, x_4 , и x_5 .

Верхняя часть лица:

Центр:

$$\left(0, \frac{x_1 + x_3}{2}\right)$$

Большая полуось:

$$x_4$$

Меньшая полуось:

$$\frac{x_1 - x_3}{2}$$

Ввод в GeoGebra:

- $A = (0, \frac{x_1+x_3}{2} + \frac{\sqrt{4*x_4^2-(x_1-x_3)^2}}{2})$ - координата первого фокуса эллипса
- $B = (0, \frac{x_1+x_3}{2} - \frac{\sqrt{4*x_4^2-(x_1-x_3)^2}}{2})$ - координата второго фокуса эллипса
- $Ellipse[A, B, x_4]$

$\frac{x_1+x_3}{2}$ - это среднее арифметическое между x_1 и x_3 , которое находит середину между верхней и нижней границами верхнего эллипса.

$\sqrt{4 * x_4^2 - (x_1 - x_3)^2}$ - это выражение определяет расстояние между двумя фокусами эллипса, где x_4 представляет эксцентриситет верхней части лица.

Нижняя часть лица:

1 Центр:

$$(0, \frac{-x_1 + x_2 + x_3}{2})$$

2 Большая полуось:

$$x_5$$

3 Меньшая полуось:

$$\frac{x_1 + x_2 + x_3}{2}$$

Ввод в GeoGebra:

- $C = (0, \frac{(-x_1+x_2+x_3)}{2} + \frac{\sqrt{4*x_5^2-(x_1+x_2+x_3)^2}}{2})$ - координата первого фокуса

- $D = (0, \frac{(-x_1+x_2+x_3)}{2} - \frac{\sqrt{4*x_5^2-(x_1+x_2+x_3)^2}}{2})$ - координата второго фокуса
- $Ellipse[C, D, x_5]$

Глаза и зрачки. Для создания эллипсов глаз и зрачков, были определены центры, полуоси и углы поворота, используя соответствующие переменные модели. В GeoGebra есть команда Rotate, чтобы повернуть фигуру вокруг точки. В случае эллипса, необходимо повернуть его вокруг его центра.

Левый и правый глаза:

1 Центр:

$$(\frac{-x_{11} - x_{14}}{2}, x_{10}) \text{ и } (\frac{x_{11} + x_{14}}{2}, x_{10})$$

2 Большая полуось:

$$x_{14}$$

3 Меньшая полуось:

$$x_{13} * x_{14}$$

Ввод в GeoGebra:

- $E = (\frac{-x_{11}-x_{14}}{2} + \sqrt{x_{14}^2 - (x_{13} * x_{14})^2}, x_{10})$ - координата первого фокуса левого глаза
- $F = (\frac{-x_{11}-x_{14}}{2} - \sqrt{x_{14}^2 - (x_{13} * x_{14})^2}, x_{10})$ - координата второго фокуса
- $AngleLeftEye = x_{12} * \frac{180}{\pi}$ - преобразование радиан в градусы
- $LeftEye = Rotate[Ellipse[E, F, x_{14}], Angle, (\frac{-x_{11}-x_{14}}{2}, x_{10})]$
- $G = (\frac{x_{11}+x_{14}}{2} + \sqrt{x_{14}^2 - (x_{13} * x_{14})^2}, x_{10})$ - координата первого фокуса правого глаза

- $H = (\frac{x_{11}+x_{14}}{2} - \sqrt{x_{14}^2 - (x_{13} * x_{14})^2}, x_{10})$ - координата второго фокуса
- $RightEye = Rotate[Ellipse[G, H, x_{14}], Angle, (\frac{-x_{11}-x_{14}}{2}, x_{10})]$

Левый и правый зрачки:

1 Центр левого и правого:

$$(-x_{11} - \frac{x_{14}}{2}, x_{10}) \text{ и } (x_{11} + \frac{x_{14}}{2}, x_{10})$$

2 Радиус:

$$R = 0.05$$

Ввод в GeoGebra:

- $Circle[(-x_{11} - \frac{x_{14}}{2} - x_{15} * \frac{x_{14}}{2}, x_{10}), 0.05]$

Рот, нос и брови. Для создания рта, носа и бровей были использованы различные команды GeoGebra для создания линий и кривых, с учетом спецификаций каждой части лица, определённых переменными модели лица Чернова.

Нос:

- $Segment[(0, \frac{-x_6}{2}), (0, \frac{x_6}{2})]$

Левая бровь:

- $StartLeftEyebrow = (\frac{-x_{11}-x_{14}}{2} - \frac{x_{14}*x_{18}}{2}, x_{10} + x_{13} * x_{14} * (x_{16} + x_{17}))$
- $EndLeftEyebrow = (\frac{-x_{11}-x_{14}}{2} + \frac{x_{14}*x_{18}}{2}, x_{10} + x_{13} * x_{14} * (x_{16} - x_{17}))$
- $Segment[StartLeftEyebrow, EndLeftEyebrow]$

Правая бровь:

- $StartRightEyebrow = (\frac{x_{11}+x_{14}}{2} + \frac{x_{14}*x_{18}}{2}, x_{10} + x_{13} * x_{14} * (x_{16} + x_{17}))$
- $EndRightEyebrow = (\frac{x_{11}+x_{14}}{2} - \frac{x_{14}*x_{18}}{2}, x_{10} + x_{13} * x_{14} * (x_{16} - x_{17}))$
- $Segment[StartRightEyebrow, EndRightEyebrow]$

Рот:

1 Центр:

$$Center = (0, -x_7 + \frac{0.5}{x_8})$$

2 Радиус:

$$R = \frac{1}{x_8}$$

3 Углы:

$$StartAngle = 270 - \arctan(x_8 x_9) * \frac{180}{\pi}$$

$$EndAngle = 270 + \arctan(x_8 x_9) * \frac{180}{\pi}$$

Ввод в GeoGebra:

- $StartPoint = (R * \cos(StartAngle), -x_7 + \frac{0.5}{x_8} + R * \sin(StartAngle))$
- $EndPoint = (R * \cos(EndAngle), -x_7 + \frac{0.5}{x_8} + R * \sin(EndAngle))$
- $CircularArc[Center, StartPoint, EndPoint]$

3.3 Визуализация в GeoGebra

В процессе визуализации модели лиц Чернова в GeoGebra были получены различные изображения, демонстрирующие множество экспрессий и эмоций. Ниже представлены изображения [3.2](#), [3.3](#), [3.4](#), иллюстрирующие различные вариации лиц, созданные с использованием разных параметров.

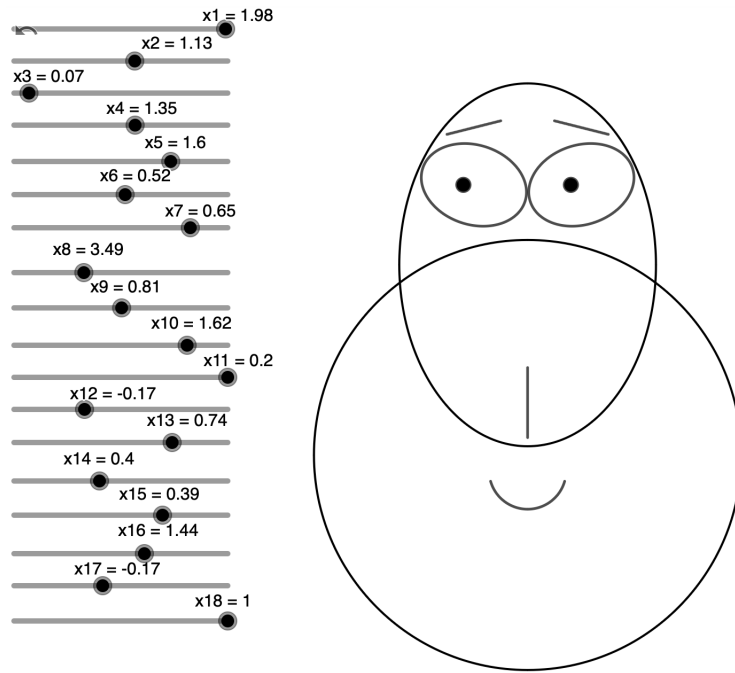


Рис. 3.2: Для создания изображения улыбающегося лица были выбраны следующие параметры: x_7 и x_8 (позиция и кривизна рта) были настроены так, чтобы подчеркнуть улыбающийся эффект, а параметры глаз x_{11} , x_{13} , x_{14} и x_{15} использовались для создания более открытого и дружелюбного взгляда

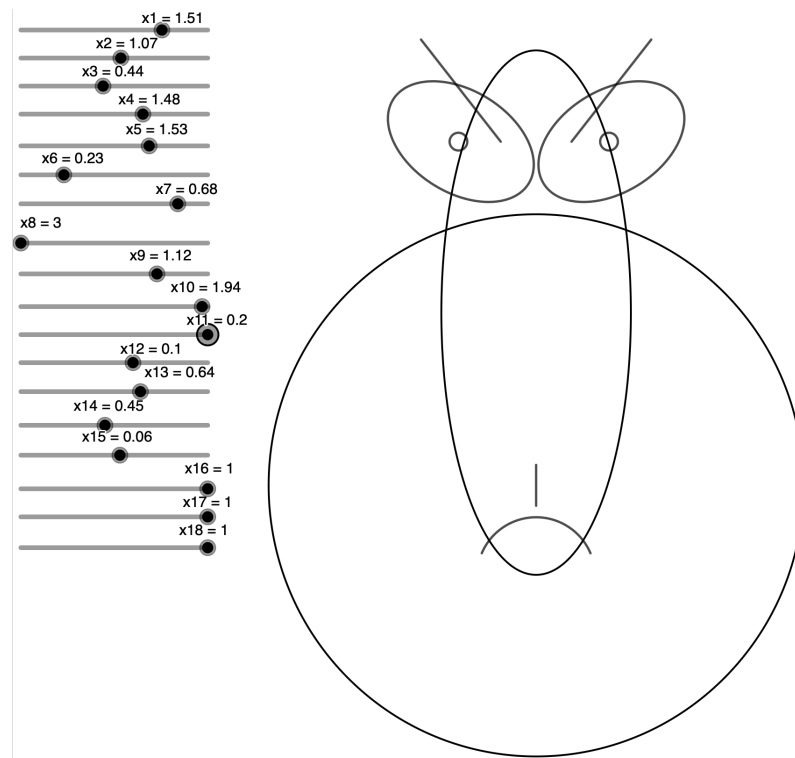


Рис. 3.3: Для создания злого лица параметры x_7 , x_8 (позиция и кривизна рта) и параметры глаз были адаптированы для того чтобы создать более агрессивное и настороженное выражение лица.

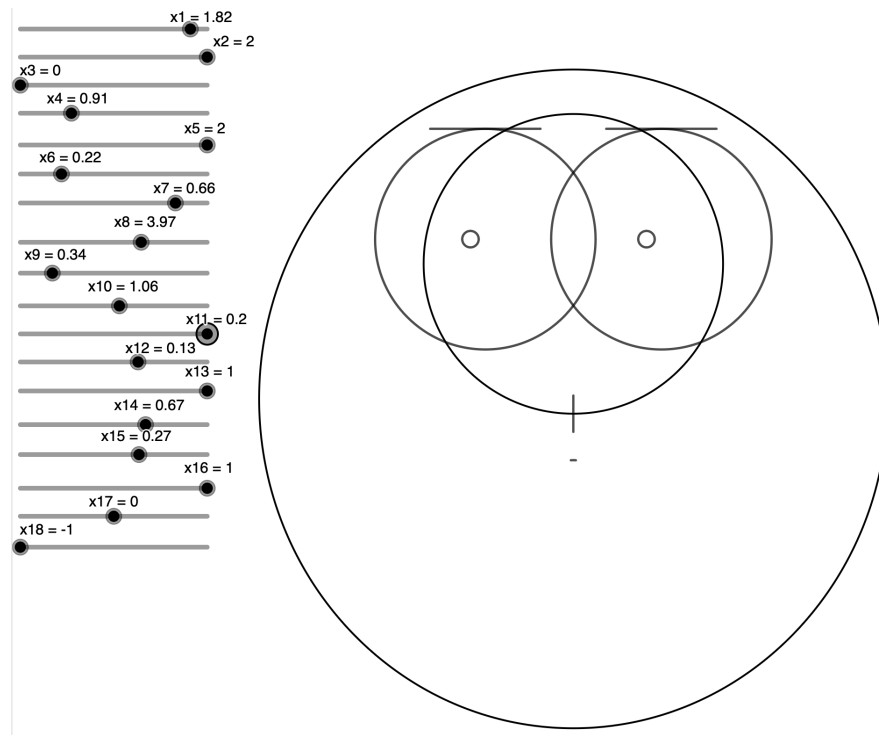


Рис. 3.4: Для создания испуганного лица были изменены параметры, связанные с глазами и ртом, чтобы подчеркнуть широко раскрытые глаза и рот.

Полученные изображения демонстрируют, насколько разнообразными и выразительными могут быть лица, созданные с использованием модели Чернова. Видно, что изменение отдельных параметров может существенно влиять на общее выражение лица, позволяя создавать портреты с различными эмоциями и характерами.

3.4 Реализация модели в Python

Код на Python был написан с целью визуализировать модель лица Чернова. Он использует библиотеку `matplotlib` и `numpy` для рисования и визуализации различных частей лица, таких как глаза, рот, нос, и брови, в соответствии с заданными параметрами модели. Каждая функция рисует часть лица с использованием параметров, которые определяют, например, размер, форму и положение. Параметры модифицируются в соответствии с константами, такими как `UPPER_FACE_MULTIPLIER` и `EYES_SHAPE_ADDITION`,

чтобы лицо было больше похоже на человеческое.

3.4.1 Основные функции

- `adjust_parameters(params)`: модификация входных параметров для обеспечения разнообразия внешности лиц. Это делается путём умножения, деления или сложения каждого параметра на определенную константу.
- `compose_face(ax, *params)`: комбинирование всех функций для создания полного изображения лица. Функция принимает объект оси и параметры лица, модифицирует параметры и вызывает другие функции для рисования частей лица на оси.
- `create_ellipse(ax, center, width, height)`: рисование эллипса для верхней и нижней частей лица. Функция принимает объект оси, координаты центра эллипса, его ширину и высоту.
- `create_mouth(ax, vertical_position, curvature, width)`: рисование рта с заданной вертикальной позицией, кривизной и шириной.
- `create_eyes(ax, vertical_position, separation, slant, eccentricity, size)`: рисование глаз и зрачков.
- `create_eyebrows(ax, vertical_position, separation, slant, eccentricity, size, brow_vertical_position, brow_slant, brow_size)`: рисование бровей на основе заданных параметров.

Пример использования кода:

```
if __name__ == "__main__":
    fig = plt.figure(figsize=(20, 20))
    for i in range(25):
        ax = fig.add_subplot(5, 5, i + 1, aspect='equal')
```

```

compose_face(ax, 0.9, *np.random.rand(17))
ax.axis([-1.2, 1.2, -1.2, 1.2])
ax.set_xticks([])
ax.set_yticks([])
fig.subplots_adjust(hspace=0, wspace=0)
plt.savefig('refactored_faces.png', bbox_inches='tight')
plt.show()

```

Код создаёт фигуру, на которой рисуются 25 лиц со случайными параметрами. Затем всё это сохраняется в файл 'refactored_faces.png' и отображается на экране. На рисунке 3.5 изображён вывод после исполнения кода.

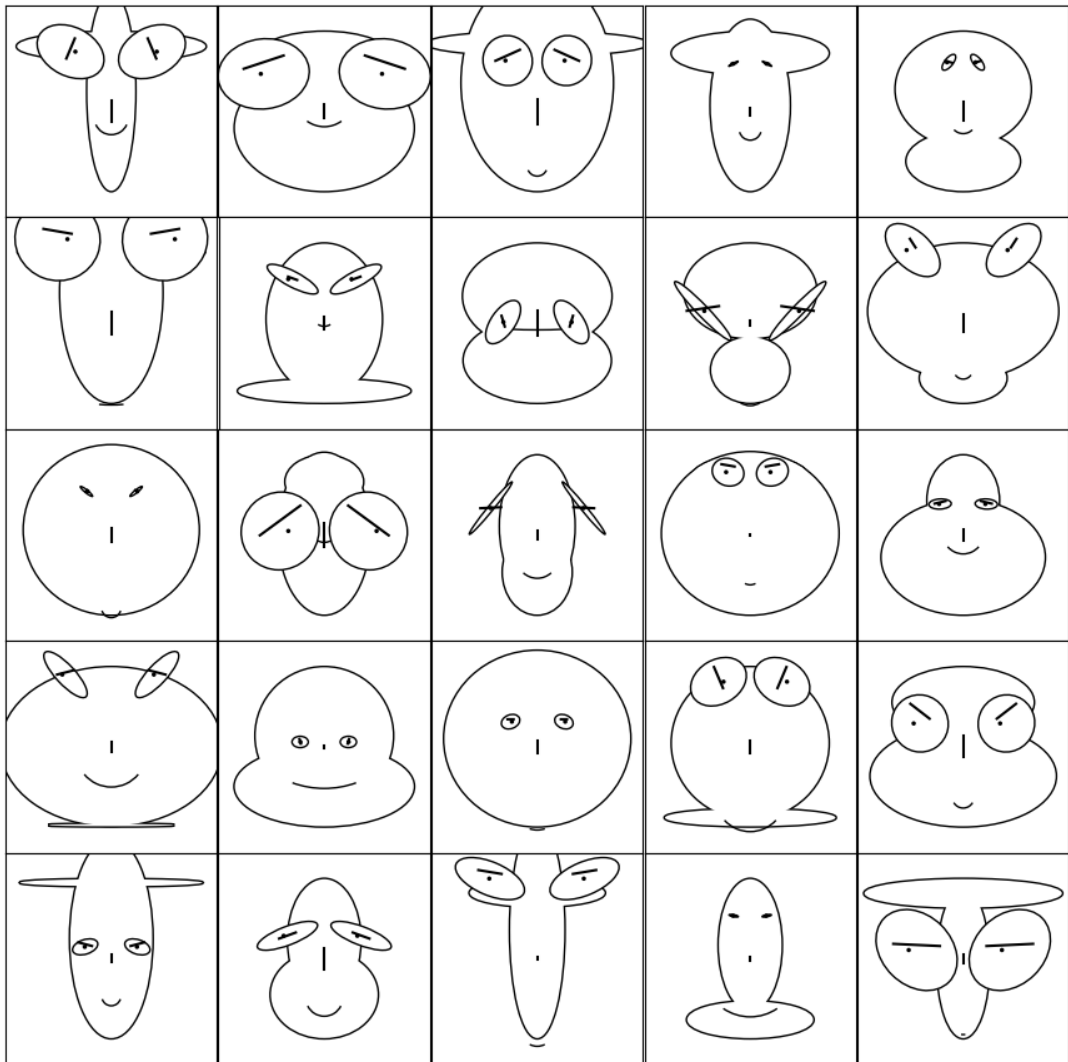


Рис. 3.5: Пример визуализации с использованием разработанного Python-кода. Параметры для отдельных элементов лица были случайно сгенерированы, демонстрируя разнообразие возможных внешностей ([репозиторий](#)).

Детальный код можно найти в репозитории по [ссылке](#).

4 Визуализация датасета Iris

В этом разделе наглядно показано, как использовать разработанную библиотеку с реальными данными. Использовались данные из Iris dataset, чтобы проиллюстрировать, как можно использовать библиотеку для визуализации.

4.1 Описание датасета Iris

Датасет Iris, возможно, является одним из самых известных датасетов в области машинного обучения. Он содержит 150 образцов ирисов 4.1, каждый из которых принадлежит одному из трёх видов: *setosa*, *versicolor* или *virginica*.

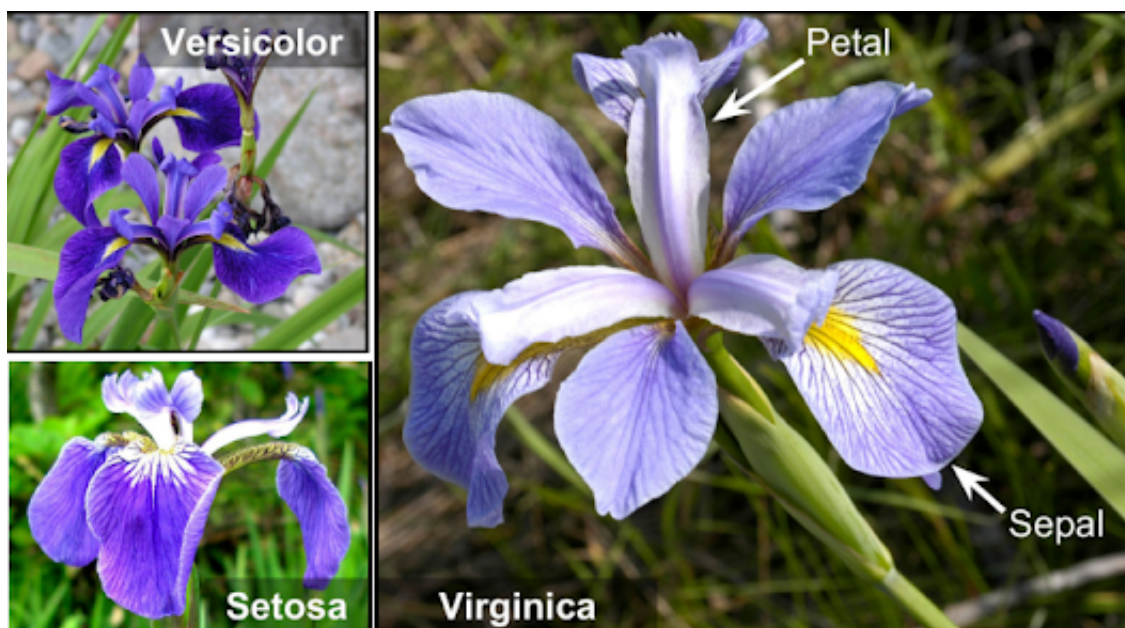


Рис. 4.1: Ирисы трёх разных видов. [18]

Для каждого образца измеряются следующие четыре характеристики:

- Длина чашелистика (sepal length)
- Ширина чашелистика (sepal width)
- Длина лепестка (petal length)

- Ширина лепестка (petal width)

Эти характеристики используются для классификации ирисов и их последующей визуализации с использованием лиц Чернова.

4.2 Использование библиотеки для визуализации

Выбор переменных

В данном случае, были выбраны следующие четыре переменные для представления числовых характеристик ирисов в лицах Чернова:

- x1 - высота верхней части лица (height of upper face)
- x4 - ширина верхней части лица (width of upper face)
- x7 - вертикальное положение рта (vertical position of mouth)
- x14 - размер глаз (size of eyes)

Выбор именно этих переменных обусловлен их способностью максимально наглядно и разнообразно представить различия между классами ирисов. Каждая из этих переменных способна визуализировать уникальные черты лица, что улучшает интерпретируемость визуализации. А для остальных переменных использовались константные значения.

Визуализация

Были созданы подграфики и нарисованы пиктограммы для каждого класса ирисов с использованием подготовленных данных. Каждый класс представлен разным цветом: 'sky blue' для setosa, 'salmon' для versicolor и 'pale green' для virginica.

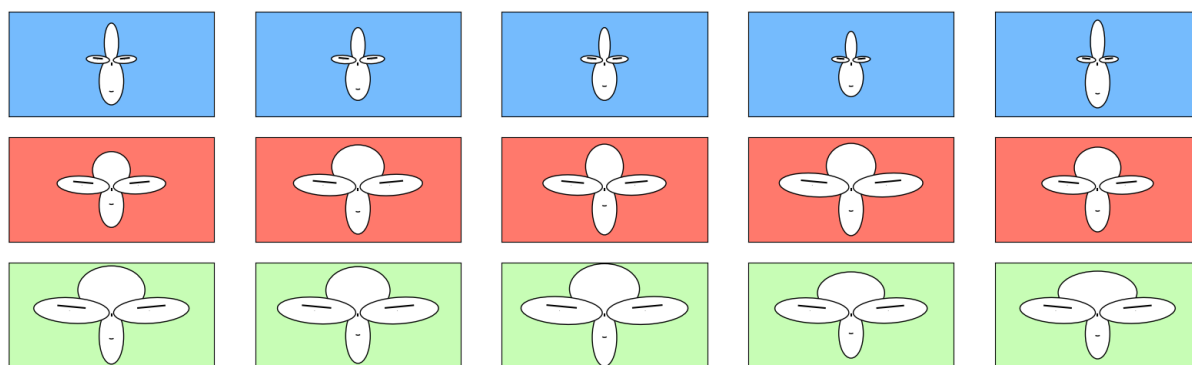


Рис. 4.2: Лица Чернова, представляющие разные виды ирисов ([репозиторий](#)).

4.3 Анализ Результатов

После применения библиотеки для визуализации данных Iris с использованием лиц Чернова, результаты были проанализированы с целью определения характерных черт каждого класса ирисов. Как видно из рисунка 4.2 первая группа сильно отличается от второй и третьей размерами глаз, верхней части лица. Второй и третий классы имеют уже меньше различий, но имеют незначительные различия в ширине верхней части лица и размере глаз.

Таким образом, были выявлены характерные черты лиц разных классов ирисов. Внимательный анализ этих пиктограмм помог выявить уникальные особенности каждой группы и дал понимание о различиях между классами на графическом уровне.

5 Заключение

Подведём итоги данного исследовательского проекта.

5.1 Результаты

В процессе выполнения данной работы была разработана и реализована модель для создания визуализации лиц по методу Чернова. Проведённый анализ существующих методик визуализации дал возможность создать оптимизированный и эффективный метод, который был воплощён в программный код на Python. Любой пользователь может воспользоваться библиотекой по [ссылке](#) на репозиторий.

В результате разработки и реализации были достигнуты следующие результаты:

- 1 Разработка модели:** Модель визуализации лиц Чернова была тщательно разработана с использованием методов математического моделирования, с фокусом на точность воспроизведения различных лиц с разными характеристиками и эмоциями.
- 2 Реализация в коде:** Код, написанный на Python с использованием библиотек `matplotlib` и `numpy`, позволяет легко модифицировать и адаптировать модель для разнообразных задач, благодаря его гибкости и масштабируемости.
- 3 Демонстрация на реальных данных:** Модель прошла тестирование и визуализировала Iris dataset, что подтвердило её эффективность и работоспособность.
- 4 Оформление в виде библиотеки:** Код был структурирован в виде удобной библиотеки, сделанной доступной для сообщества разработчиков.

5.2 Будущее развитие

Возможные улучшения и доработки

- **Расширение функционала:** Существует потенциал для расширения возможностей библиотеки, включая добавление новых методов визуализации и улучшение существующих.
- **Оптимизация кода:** Есть возможность для дальнейшей оптимизации кода для повышения его производительности и эффективности.
- **Интеграция с другими инструментами:** Модель можно интегрировать с другими инструментами анализа данных и машинного обучения, расширяя тем самым область её применения.

Применимость в реальных задачах

- **Образование и научные исследования:** Модель может служить ценным инструментом для обучения и исследований в области визуализации данных и многомерных пространств.
- **Анализ данных:** Визуализация лиц Чернова может найти применение в различных сферах для анализа многомерных данных, таких как биоинформатика, финансовый анализ и социология.
- **Развлекательная индустрия:** Возможности модели могут быть использованы в игровой разработке и других развлекательных областях для создания уникальных визуализаций и интерфейсов.

Заключительные замечания

Реализованная модель лиц Чернова открывает новые горизонты для визуализации многомерных данных, предлагая высокую степень гибкости

и адаптивности, что позволяет её использовать в различных контекстах и областях. Эта работа, вероятно, окажет положительное влияние на исследователей и практиков в области анализа данных, машинного обучения и визуализации данных.

6 Календарный план работы над проектом

- **21.01.23** - Зачисление на проект, изучение материалов по теме работы, изучение прикладной области применения.
- **14.02.23** - Изучение научных работ - источников. Разработка подробного плана работы. Написание отчета для КТ-1.
- **06.03.23** - Реализация геометрической модели лиц Чернова в системе GeoGebra.
- **20.03.23** - Реализация разработанной модели в программном коде на языке Python.
- **20.04.23** - Оформление кода в виде библиотеки.
- **29.04.23** - Написание документации и демонстрация применения разработанной библиотеки на реальных статистических данных.
- **15.05.23** - Написание отчета.
- **22.05.23** - Подготовка презентации.
- **30.05.23** - Перепроверка всего проекта, согласование отчета с руководителем.
- **Июнь 2023** - Защита проекта.

Список используемой литературы

1. Addelman S. Orthogonal main-effect plans for asymmetrical factorial experiments //Technometrics. – 1962. – Т. 4. – №. 1. – С. 21-46.
2. Wang P. C. C. (ed.). Graphical representation of multivariate data. – Elsevier, 2014.
3. Chernoff H., Rizvi M. H. Effect on classification error of random permutations of features in representing multivariate data by faces //Journal of the American Statistical Association. – 1975. – Т. 70. – №. 351a. – С. 548-554.
4. Aghababian R. et al. Selection of emergency medicine residents //Annals of emergency medicine. – 1993. – Т. 22. – №. 11. – С. 1753-1761.
5. Chernoff H. The use of faces to represent points in k-dimensional space graphically //Journal of the American statistical Association. – 1973. – Т. 68. – №. 342. – С. 361-368.
6. RAHU M. Graphical representation of cancer incidence data: Chernoff faces //International journal of epidemiology. – 1989. – Т. 18. – №. 4. – С. 763-767.
7. Flury B., Riedwyl H. Graphical representation of multivariate data by means of asymmetrical faces //Journal of the American Statistical Association. – 1981. – Т. 76. – №. 376. – С. 757-765.
8. Kabulov B. T. A method for constructing chernoff faces oriented toward interval estimates of the parameters //Soviet journal of computer and systems sciences. – 1992. – Т. 30. – №. 3. – С. 94-97.
9. Osadchaya I. A., Berestneva O. G., Nemerov Y. V. ANALYSIS OF MULTIDIMENSIONAL MEDICAL DATA USING PICTOGRAPHICS

- «CHERNOFF FACES» //Bulletin of Siberian medicine. – 2014. – Т. 13. – №. 4. – С. 89-93.
10. Reyes Nuñez J. J. Ideas for the use of Chernoff faces in school cartography //World's Geo-Spatial Solut. – 2009.
 11. Aghababian R. et al. Selection of emergency medicine residents //Annals of emergency medicine. – 1993. – Т. 22. – №. 11. – С. 1753-1761.
 12. Визуализация. — Текст : электронный // Википедия : [сайт]. — URL: <https://ru.wikipedia.org/wiki/Визуализация> (дата обращения: 10.02.2023).
 13. Визуализация данных. — Текст : электронный // Яндекс.Практикум : [сайт]. — URL: <https://practicum.yandex.ru/blog/vizualizaciya-dannyh/> (дата обращения: 10.02.2023).
 14. Дашборд. — Текст : электронный // Яндекс.Практикум : [сайт]. — URL: <https://practicum.yandex.ru/blog/vizualizaciya-dannyh/> (дата обращения: 10.02.2023).
 15. Пиктограмма. — Текст : электронный // Википедия : [сайт]. — URL: <https://en.wikipedia.org/wiki/Pictogram> (дата обращения: 10.02.2023).
 16. Chernoff faces. — Текст : электронный // esri : [сайт]. — URL: <https://www.esri.com/content/dam/esrisites/en-us/esri-press/pdfs/cartography-definitive-guide-making-maps-sample-chapter.pdf> (дата обращения: 15.02.2023).
 17. Chernoff-Gesichter. — Текст : электронный // WikipediA : [сайт]. — URL: <https://de.wikipedia.org/wiki/Chernoff-Gesichter> (дата обращения: 20.08.2023).

18. Iris Dataset Classification with Python. — Текст : электронный // PyCodeMates : [сайт]. — URL: <https://www.pycodemates.com/2022/05/iris-dataset-classification-with-python.html> (дата обращения: 29.08.2023).