

МЕТОД ОПРЕДЕЛЕНИЯ ЛИЦЕВЫХ ДВИЖЕНИЙ ЧЕЛОВЕКА НА ФОТОГРАФИИ

Спасенов Иван Владимирович

студент, Московский государственный технический университет им. Н. Э.

Баумана,

Россия, г. Москва

METHOD FOR DETERMINING FACIAL ACTION UNITS IN THE IMAGE

Spasenov Ivan Vladimirovich

student, Bauman Moscow State Technical University,

Russia, Moscow

АННОТАЦИЯ

Целью работы является разработка метода определения лицевых движений на фотографии. В работе проводился обзор существующих методов решения задачи. Был разработан и реализован метод определения лицевых движений. Были рассмотрены преимущества и недостатки разработанного метода и предложены пути дальнейшего развития. Разработано программное обеспечение на основе предложенного метода.

ABSTRACT

The purpose of this work is to develop a method for determining facial action units in the image. The paper reviewed existing methods for solving the problem. A method for determining facial action units was developed and implemented. The advantages and disadvantages of the developed method were considered and ways of further development were suggested. The software is developed based on the proposed method.

Ключевые слова: лицевые движения, анализ изображения, машинное обучение, дерево решений, бустинг, лэндмарки.

Keywords: facial action units, image analysis, machine learning, decision tree, boosting, facial landmarks.

Введение

Под движениями лица подразумеваются определенные форма и положение частей лица, выражающие внутреннее душевное состояние [1]. Определение лицевых движений является важной задачей, так как они позволяют сделать вывод о психофизическом состоянии, человека [2]. Так, анализ лицевых движений используется, в частности, для определения эмоций, оценки степени депрессии [11] и измерения боли пациентов, которые не в состоянии говорить.

Целью проекта является разработка метода для определения лицевых движений на изображении лица человека.

Система кодирования лицевых движений

В качестве системы кодирования лицевых движений использована система, изначально разработанная Экманом и Фризенем в 1978 году [2].

На рисунке 1 показаны примеры некоторых лицевых движений.































AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Поднятые внутренние части бровей	Поднятые внешние части бровей	Опушенные брови	Поднятые внутренние части бровей	Поднятые щеки	Натянутое веко
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Опушенное верхнее веко	Узкий разрез глаз	Глаза закрыты	Прищур	Моргание	Подмигивание
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Сморщенный нос	Поднятая верхняя губа	Углубленная носогубная складка	Поднятые уголки губ	Остро поднятые уголки губ	Ямочка на щеке
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Опушенные уголки губ	Опушенная нижняя губа	Поднятый подбородок	Сморщенные губы	Растянутые губы	Губы воронкой
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Натянутые губы	Сжатые губы	Губы разведены	Челюсть опущена	Рот широко открыт	Вытянутые губы

Рисунок 1. Примеры лицевых движений

Обзор существующих методов

- Метод, основанный на нейронной сети JAA-Net.

Метод предложен Шао и Лио в 2018 г. [5] Основу метода составляют четыре модуля: иерархическое и многомасштабное обучение, модуль для получения ключевых точек лица, глобальное обучение представлений и адаптивное обучение внимания, рисунок 2. Иерархическое и многомасштабное обучение разработано, как основа JAA-Net, оно извлекает характеристики каждого локального региона изображения в разных масштабах. Модуль, получающий лицевые точки, помимо генерации лэндмарков используется для генерации начальных карт внимания, которые используются для обнаружения лицевых движений. Модуль глобального обучения представлений призван охватить особенности структуры и текстуры всего лица. Адаптивное обучение внимания разработано как центральная часть для обнаружения лицевых движений, этот модуль изучает карту внимания каждого лицевого движения.

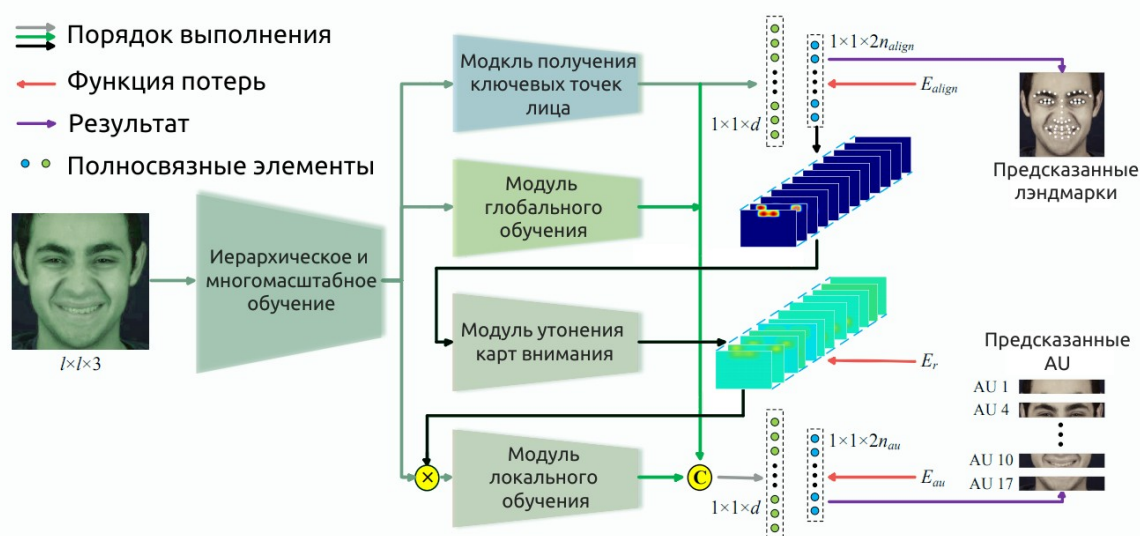


Рисунок 2. Архитектура JAA-Net

- Метод, основанный на нейронной сети ARL.

Метод предложен Шао и Лио в 2019 г. [6]. Структура ARL состоит из трех частей: иерархического и многомасштабного обучения, обучения пространственному вниманию на уровне каналов и обучения выявлению связей

на уровне пикселей. Модуль иерархического и многомасштабного обучения областей направлен на извлечение многомасштабных представлений для адаптации к различным лицевым движениям с различными размерами.

- Метод, основанный на нейронной сети DRML.

Метод предложен Шао и Чу в 2016 г. [7]. Входными данными является изображение лица в RGB формате, которое затем передается на сверточный слой (conv1) с 32 фильтрами размером $11 \times 11 \times 3$. Слой conv1 генерирует 32 карты объектов, которые подаются в слой областей (region2). Далее следует слой пуллинга по максимальному значению (pool3) 2×2 с шагом 2, отдельно для каждого канала карт объектов. Слой pool3 делает сеть более устойчивой к небольшим ошибкам, вызванным выравниванием лица. В DRML используется только один слой пуллинга по максимальному значению, чтобы не потерять слишком много пространственной информации. За слоем pool3 следуют еще четыре сверточных и два полносвязных слоя. В сети применяется сигмоидальная кросс-энтропийная функция потерь.

Сравнение методов определения мимики человека на изображении

Достоинства и недостатки рассмотренных методов приведены в таблице 1.

Таблица 1.

Достоинства и недостатки методов определения мимики человека.

Название метода	Преимущества	Недостатки
метод на основе нейронной сети JAA-Net	+ высокая точность работы (средняя F1 мера по всем AU = 60.0) + определение положения ключевых точек[5]	- сложная реализация - низкая скорость работы
метод на основе	+ точность работы выше,	- низкая скорость

нейронной сети ARL	чем у остальных рассмотренных методов (средняя F1 мера по всем AU = 61.1) [6]	работы
метод на основе нейронной сети DRML	+ высокая скорость работы	- точность работы меньше, чем у остальных рассмотренных методов (средняя F1 мера по всем AU = 48.3) [7]

Разработка метода определения лицевых движений

Для распознавания лицевых движений на фотографии в первую очередь необходимо определить на ней положение лица человека, которое будет использоваться для последующей обработки. Для этого используется метод на основе HOG [3], так как он сочетает в себе высокую скорость работы на CPU и хорошую точность при незначительных перекрытиях объектов.

Лицевые движения формируются частями лица такими как губы, брови, глаза и тд. Их форму и положение показывают ключевые точки лица - лэндмарки. Определение ключевых точек осуществляется с помощью метода на основе адаптивного бустинга деревьев решений [4], так как он обладает высокой скоростью работы на CPU и имеет приемлемую точность при незначительных поворотах головы и перекрытиях. Найденные ключевые точки поворачиваются таким образом, чтобы линия, соединяющая центры глаз, стала горизонтальной. Это необходимо, чтобы обеспечить работу алгоритма при поворотах головы в плоскости изображения. Для того чтобы метод работал с лицами разного размера одинаково хорошо, обрабатываемое лицо масштабируется таким образом, чтобы расстояние между центрами глаз стало константным значением (было выбрано значение 100 пикселей).

Первоначально по определенным ранее координатам ключевых точек глаз находятся координаты центров глаз. Затем рассчитывается угол наклона линии, соединяющей центры глаз – a , расстояние между центрами – L , а также координаты центра поворота изображения – координаты середины отрезка, соединяющего центры глаз – $(centerX, centerY)$.

Для поворота и масштабирования изображения составляется матрица поворота, с параметрами угол($angle$) = a , коэффициент масштабирования($scale$) = $100 / L$, центр поворота($center$) = $(centerX, centerY)$:

$$\begin{bmatrix} \alpha & \beta & (1 - \alpha) \cdot center.x - \beta \cdot center.y \\ -\beta & \alpha & \beta \cdot center.x + (1 - \alpha) \cdot center.y \end{bmatrix}$$

$$\begin{aligned} \alpha &= scale \cdot \cos angle, \\ \beta &= scale \cdot \sin angle \end{aligned}$$

Матрица с координатами ключевых точек умножается на полученную матрицу поворота.

В основе разрабатываемого метода лежит предположение о том, что по оценке положения ключевых точек лица относительно друг друга можно сделать вывод о наличии лицевых движений. Это возможно, так как лэндмарки способны передавать положение и форму бровей, рта, носа, глаз и челюсти (рисунок 3). Ключевые точки не учитывают образующие при лицевых движениях морщины [9], исходя из этого для классификации были выбраны лицевые движения, которые явно передаются формой и положением частей лица: AU1 (поднятые внутренние части бровей), AU2 (поднятые внешние части бровей), AU4 (опущенные брови), AU5 (поднятое верхнее веко), AU12 (поднятые уголки губ), AU15 (опущенные уголки губ), AU25 (губы разведены), AU26 (челюсть опущена). Для каждого лицевого движения из полученных значений координат лэндмарков необходимо получить вектор значений, по которому можно было бы классифицировать данный AU.

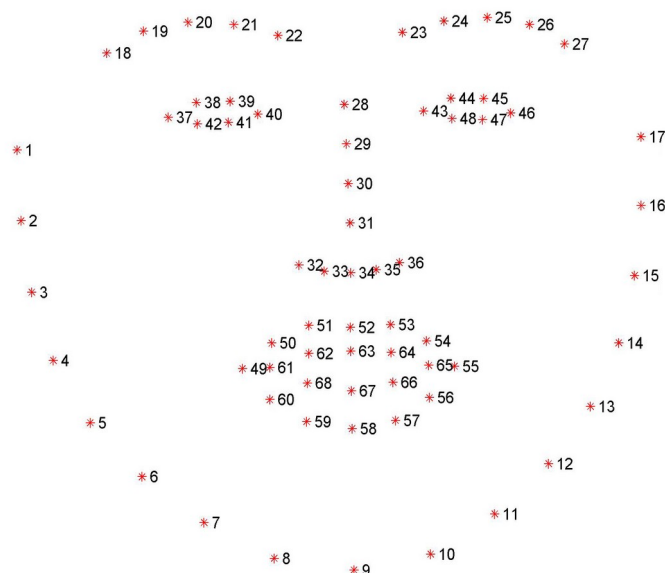


Рисунок 3. Ключевые точки лица

AU1, AU2, AU4 основаны на положении и форме бровей, как видно на примере, показанном на рисунке 4.



Рисунок 4. Пример AU1, AU2, AU4

Как следует из системы кодирования лицевых движений [2] характерными чертами для AU1 являются поднятые внутренние части бровей, для AU2 - поднятые внешние части бровей, для AU4 - опущенные брови. Исходя из этого введем в качестве признаков следующие величины: расстояния по вертикали между внешними и внутренними частями бровей (точки 18, 22 для левой 23, 27 для правой брови, рисунок 3), расстояния по вертикали между бровями (точки 18-22 и 23-27) и верхней точкой носа (точка 28) (точка выбрана, так как является неподвижной во время движения бровей), расстояние по горизонтали между бровями (точки 22 и 23).

Расстояния между лэндмарками, которые соответствуют упомянутым частям лица, рассчитывались в пикселях. Расчет производился на наборе данных DISFA [10], в таблице 2 представлено количество изображений по каждому AU, которые он содержит.

Таблица 2.

Количество изображений по каждому AU

	AU1	AU2	AU4	AU5	AU12	AU15	AU25	AU26
Кол-во	8778	7362	20857	2729	30794	7862	37673	22872

AU5 основан на положении век, как видно на примере, показанном на рисунке 5.

AU 5



Рисунок 5. Пример AU5

Как следует из системы кодирования лицевых движений [2] характерными чертами для AU5 поднятое верхнее веко относительно нижнего. Исходя из этого введем в качестве признака расстояние по вертикали между верхними и нижними веками (точки 38, 39, 44, 45 и 42, 41, 48, 47 соответственно).

AU12, AU15, AU25 основаны на положении и форме губ, как видно на примере, показанном на рисунке 6.



Рисунок 6. Пример AU12, AU15, AU25

Как следует из системы кодирования лицевых движений [2] характерной чертой для AU12 является поднятые уголки губ, для AU15 - опущенные уголки губ, для AU25 - разведенные губы. Исходя из этого для AU12 и AU15 введем в качестве признаков расстояние по вертикали между уголками губ (49, 55) и верхней губой (точки 50 - 54), а также между уголками губ и нижней губой (точки 56 - 60). Для AU25 признаком выберем расстояние между губами (точки 62-64 и 66-68 соответственно).

AU26 основан на положении челюсти, как видно на примере, показанном на рисунке 7.



Рисунок 7. Пример AU26

Как следует из системы кодирования лицевых движений [2] характерной чертой для AU26 является опущенная челюсть. Исходя из этого в качестве признаков было выбрано расстояние по вертикали между нижней челюстью (точки 6 - 12) и нижней точкой носа (точка 34). (точка выбрана, так как она является неподвижной во время движения челюсти).

Классификация лицевых движений

По полученным векторам признаков для каждого AU проводилась бинарная классификация его наличия на фото. В качестве классификатора использовался градиентный бустинг из деревьев решений, так как этот метод считается одним из наиболее эффективных с точки зрения качества классификации. В качестве модели было выбрано неглубокое решающее дерево, так как они обладают большим смещением и маленьким разбросом, они

не склонны к переобучению, а благодаря бустингу понижается смещение модели.

Для градиентного бустинга из деревьев использовалась библиотека xgboost. Параметры для бустинга подбирались в процессе кросс-валидации [8].

Параметры имели следующие значения:

- для AU1, AU2, AU4
 - высота дерева = 3
 - количество деревьев = 100
 - обучающий коэффициент = 0.1
- для AU5, AU25
 - высота дерева = 3
 - количество деревьев = 50
 - обучающий коэффициент = 0.2
- для AU12, AU15, AU26
 - высота дерева = 3
 - количество деревьев = 100
 - обучающий коэффициент = 0.2

Набор данных

В качестве набора данных использовался датасет DISFA [10], в нем присутствуют выбранные для классификации лицевые движения. Данный набор содержит 27 видео, для каждого кадра которых описаны присутствующие на нем лицевые движения - AU. В датасете описаны 12 AU: AU1 (поднятые внутренние части бровей), AU2 (поднятые внешние части бровей), AU4 (опущенные брови), AU5 (поднятое верхнее веко), AU6 (поднятые щеки), AU9 (сморщенный нос), AU12 (поднятые уголки губ), AU15 (опущенные уголки губ), AU17 (поднятый подбородок), AU20 (растянутые губы), AU25 (губы разведены), AU26 (челюсть опущена).

Датасет был подвергнут предварительной обработке. Видео были разбиты на кадры и для каждого кадра выполнен поиск лица. Для этого был использован метод на основе HOG [3]. Найденное лицо масштабировалось и

поворачивалось таким образом, чтобы линия, соединяющая глаза, была горизонтальна. Далее определялись координаты лэндмарков с помощью метода на основе бустинга деревьев решений [4]. Полученные лэндмарки сохранялись в файлы.

Пример работы метода

На рисунке 8, представлен пример работы программы, основанной на разработанном методе.

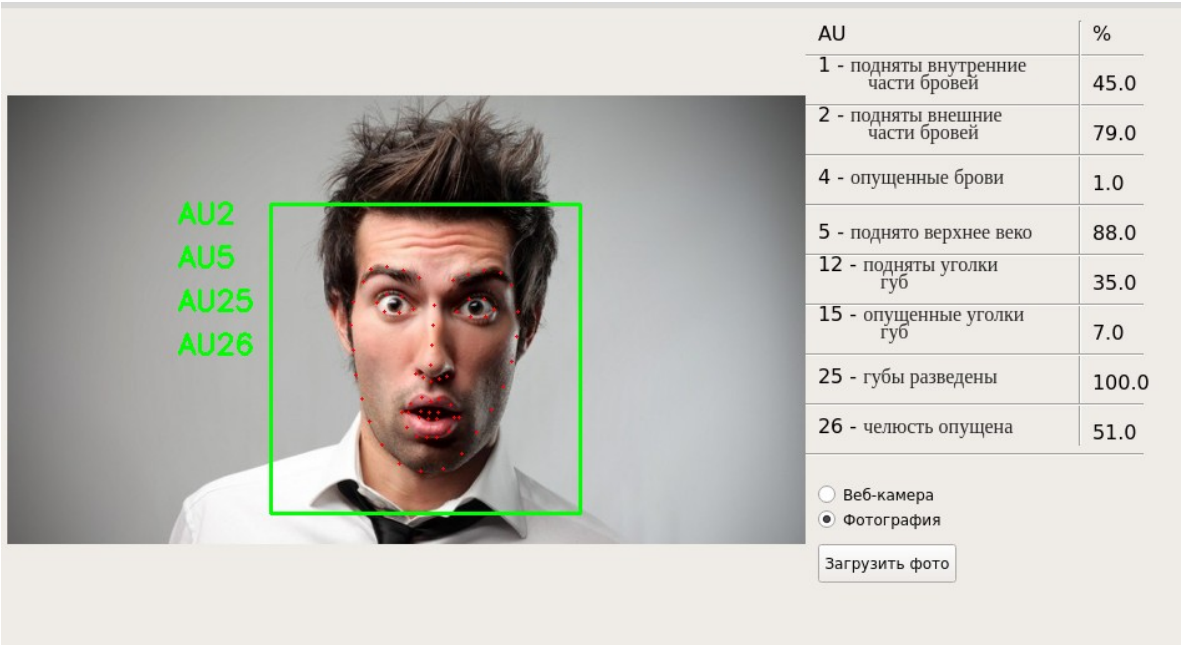


Рисунок 8. Обработка изображения

Сравнительные результаты работы метода определения лицевых движений

В качестве метрик оценивания использовались точность и F1 мера. В таблице 3 представлено сравнение метрик предложенного метода и других рассмотренных методов на датасете DISFA.

Таблица 3.

Сравнительные результаты метрик методов определения лицевых движений

AU	F1 мера		Точность (%)
----	---------	--	--------------

	JAA-Net	ARL	DRML	предложенный метод	JAA-Net	ARL	DRML	предложенный метод
1	43.7	43.9	17.3	23.6	93.4	92.1	53.3	65.5
2	46.2	42.1	17.7	25.1	96.1	92.7	53.2	66.3
4	56.0	63.6	37.4	36.2	86.9	88.5	60.0	57.8
5	НД	НД	НД	28.4	НД	НД	НД	68.2
12	69.6	76.2	37.7	50.2	91.2	93.9	54.6	75.4
15	НД	НД	НД	41.3	НД	НД	НД	64.5
25	88.3	95.2	38.5	62.2	93.4	97.3	45.6	65.1
26	58.4	66.8	20.1	31.8	93.2	94.3	45.3	58.4

Как видно из таблицы 3, предложенный метод показывает лучшие результаты по сравнению с DRML, однако все же худшие, чем у ARL и JAA-Net.

Сравнительные результаты времени выполнения метода определения лицевых движений

В таблице 4 представлено сравнение времени выполнения предложенного метода и остальных рассмотренных методов. Замеры производились на CPU Intel(R) Core(TM) i5-8250U CPU @1.60GHz. В качестве результата бралось усредненное значение для 100 измерений времени выполнения алгоритма для фотографии 200x200 пикселей.

Таблица 4.

Сравнительные результаты времени выполнения методов определения лицевых движений

	JAA-	ARL	DRML	предложенный метод
--	------	-----	------	--------------------

	Net			
время выполнения (сек.)	2.62	5.63	1.51	0.43

Как видно из таблицы 4, предложенный метод показывает значительно лучшие результаты времени выполнения по сравнению с остальными рассмотренными методами.

Заключение

В ходе данной работы были проанализированы существующие методы определения положения лица. Был разработан метод определения лицевых движений AU1 (поднятые внутренние части бровей), AU2 (поднятые внешние части бровей), AU4 (опущенные брови), AU5 (поднятое верхнее веко), AU12 (поднятые уголки губ), AU15 (опущенные уголки губ), AU25 (губы разведены), AU26 (челюсть опущена) на изображении. Были проведены сравнения метрик качества и скорости выполнения разработанного метода определения лицевых движений с существующими алгоритмами, которые показали, что разработанный метод имеет лучшие показания по скорости выполнения среди всех рассмотренных методов, а также лучшие результаты точности и F1 меры по сравнению с алгоритмом DRML, уступив при этом ARL и JAA-Net.

В качестве перспектив дальнейшего развития работы можно предложить увеличение количества классифицируемых лицевых движений, а также использование трехмерной вместо двумерной карты ключевых точек, что позволит увеличить точность метода при вращении головы.

Список литературы:

1. Толковый словарь Ожегова [Электронный ресурс] – Режим доступа: <https://dic.academic.ru/dic.nsf/ogegova/107398> (Дата обращения: 07.05.2020)
2. Ekman P. Facial Action Coding System: A Technique for the Measurement of Facial Movement. / Ekman P., Friesen W. // CPP - 1978
3. Dalal N. Histograms of Oriented Gradients for Human Detection / Dalal N., Triggs B. // IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) – 2005 - С. 2.
4. Kazemi V. One Millisecond Face Alignment with an Ensemble of Regression Trees / Kazemi V., Sullivan J. // Computer Vision and Pattern Recognition - 2014
5. Zhiwen Shao, Deep Adaptive Attention for Joint Facial Action Unit Detection and Face Alignment / Zhiwen Shao, Zhilei Liu, Jianfei Cai, Lizhuang Ma // ECCV - 2018
6. Zhiwen Shao, Facial Action Unit Detection Using Attention and Relation Learning / Zhiwen Shao, Zhilei Liu, Jianfei Cai, Senior Member, Yunsheng Wu, and Lizhuang Ma // IEEE Transactions on Affective Computing - 2019
7. Vítor Albiero, Deep Region and Multi-label Learning for Facial Action Unit Detection / Vítor Albiero, Luciano Silva, Olga Regina Pereira Bellon// IEEE International Conference on Image Processing (ICIP) - 2018
8. Кросс валидация [Электронный ресурс] – Режим доступа: <https://neerc.ifmo.ru/wiki/index.php?title=Кросс-валидация> (Дата обращения: 07.05.2020)
9. Irenäus Eibl-Eibesfeldt, Human Ethology – Германия, Aldine de Gruyter, 1989
10. Mavadati, S.M. Disfa: A spontaneous facial action intensity database. / Mavadati, S.M., Mahoor, M.H., Bartlett, K., Trinh, P., Cohn, J.F. // IEEE Transactions on Affective Computing – 2013
11. Reed L. Impact of depression on response to comedy: A dynamic facial coding analysis / Reed L. I., Sayette M. A., Cohn J. F. // Journal of abnormal psychology: journal. — 2007