

Қазақстан Республикасы Тәуелсіздігінің 30 жылдығына арналған «Сейфуллин оқулары – 17: «Қазіргі аграрлық ғылым: цифрлық трансформация» атты халықаралық ғылыми – тәжірибелік конференцияға материалдар = Материалы международной научно – теоретической конференции «Сейфуллинские чтения – 17: «Современная аграрная наука: цифровая трансформация», посвященной 30 – летию Независимости Республики Казахстан.- 2021.- Т.1, Ч.4 - С.151-155

## **ИССЛЕДОВАНИЕ АЛГОРИТМОВ И РЕАЛИЗАЦИЯ МЕТОДОВ РАСПОЗНАВАНИЯ ЭМОЦИЙ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ**

*Адикова А.К.*

Распознавание эмоций существует уже давно. Но в современном развивающемся мире, просто распознать лицо человека оказалось не достаточным. Распознавание эмоций в последнее время привлекает к себе постоянное внимание исследователей. Эмоции не только играют решающую роль во взаимодействии, принятии решений и когнитивных процессах, но также влияют на точность нашей памяти о некоторых событиях. В взаимоотношениях между собой люди естественным образом выражают и распознают свои эмоции с помощью различных модальностей, включая речь, семантику, лицевые сигналы и физические движения [1].

Обнаружение человеческих эмоций - необходимость настоящего времени, чтобы системы искусственного интеллекта могли имитировать и измерять реакции по лицу. Это может быть полезно для принятия решений, связанные с безопасностью. Распознавание эмоций на изображениях или видео – одна из легких задач для человеческого глаза, но оказывается очень сложной для машин и требует множества методов обработки изображений для извлечения признаков.

Вместе с тем в последние годы автоматическое определение эмоционального состояния человека постоянно привлекает внимание исследователей из-за его широкого спектра приложений и растущего спроса во многих различных областях, таких как взаимодействие человека с роботом, безопасное вождение, настройка веб-сайтов и образование [3]. Например, система обнаружения эмоций водителя может автоматически определять эмоциональное состояние водителя и принимать соответствующие меры для обеспечения безопасности дорожного движения и здоровья человека. [2].

Человеческие эмоции можно разделить на страх, презрение, отвращение, гнев, удивление, грусть, счастье и нейтрального характера. Эти эмоции очень тонкие. Сокращения лицевых мышц очень минимальны, и выявить эти различия очень сложно, поскольку даже небольшая разница приводит к разным выражениям лица. Кроме того, выражения разных или даже одних и тех же людей могут различаться для одной и той же эмоции,

поскольку эмоции сильно зависят от контекста. Пока исследование сосредоточится только на этих областях лица, которые отображают максимум эмоций, например, вокруг рта и глаз [4], как извлекать эти жесты и классифицировать их по-прежнему остается важным вопросом. Для этих задач использовались нейронные сети и машинное обучение.

Данное исследование направлено на исследовании алгоритмов и методов распознавания с использованием машинного обучения. Рассмотрим возможности современных алгоритмов распознавания лиц, некоторым важным характеристикам лица, использования различных типов методов для определения эмоций.

Первым алгоритмом распознавания лиц была использование функций HoG на основе динамической сетки. Это детектор лиц Dlib основанный на функциях гистограммы ориентированных градиентов (HoG) – метод обнаружения объекта, основанный на информации о краях. Каждое окно обнаружения может быть определено ориентацией краев. Изображение обрезается в соответствии с интересующей областью лица. Это обрезанное лицо является окном обнаружения, которое разделено на еще меньший набор областей, называемых ячейками. В каждой ячейке для каждого пикселя вычисляется величина краевого градиента для каждого интервала ориентации, таким образом формируя локальную гистограмму ориентированных градиенты [4].

Как показано на рисунке 1, с лицевой стороны требуемая область обрезается и делится на матрицу из 8 строк x 6 столбцов. Размер пикселя может отличаться в сетке для каждой ячейки.

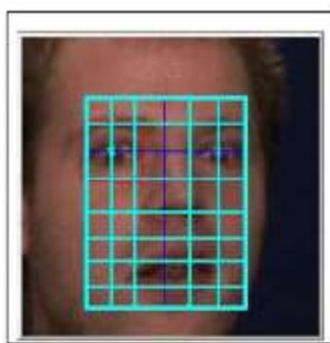


Рисунок 1: Обрезанное изображение, разделенное на ячейки по 8 строк по 6 столбцов [4]

Следующий рассмотренный алгоритм распознавания – это каскадный детектор лиц Хаара (Haar Cascade Face Detector) в OpenCV. Обнаружение лиц с использованием каскадных классификаторов Хаара - это эффективный подход, основанный на машинном обучении, который был самым используемым в течение многих лет с 2001 года, когда его предложили Пол Виола и Майкл Джонс. В этом методе каскадная функция обучается на большом количестве изображений с лицами и без лиц, а затем используется для обнаружения лиц на других изображениях. Вместе с тем за последние годы в этот подход было внесено много улучшений.

Еще один алгоритм распознавания – это детектор лиц MTCNN (многозадачная каскадная сверточная нейронная сеть) [5]. Данный подход, основанный на глубоком обучении, который использует три каскадных сверточных нейронных сети (CNN) для быстрого и точного обнаружения лиц. В MTCNN обнаружение лиц и выравнивание лиц выполняются вместе в режиме многозадачного обучения, что позволяет лучше обнаруживать невыровненные лица. Применение данного метода в сочетании с HoG является объектом исследования в статье по распознаванию групповых эмоций, которая тоже сейчас привлекает большое внимание ученых.

Также хочется остановиться на еще одном методе распознавания лиц – это детектор Виолы-Джонса. В данном детекторе изображения предварительно обрабатываются для уменьшения шума. Также в наборе данных должны быть изображения с одинаковым уровнем экспозиции, освещения и яркости. Улучшение изображения выполняется на таких изображениях с использованием методов выравнивания гистограммы. Модуль обнаружения лиц представляет собой знакомую технику Виолы-Джонса, в которой используется алгоритм Ada-boost. [4].

Вместе с тем при распознавании лиц есть особенные показатели, которые помогают правильно идентифицировать объект распознавания. Обычно такими идентифицированными признаками является углы и края. Для поиска таких углов и краев на изображении в вышеупомянутой библиотеке Open CV есть множество алгоритмов детектора признаков, такие как детектор углов Харриса. Эти детекторы особенностей учитывают множество других факторов, таких как контуры, корпус и выпуклость. Ключевые точки - это угловые точки или края, обнаруженные алгоритмом обнаружения признаков.

Можно также остановиться на детекторе лиц Deep Neural Network в OpenCV основан на платформе Single Shot Detector (SSD) с архитектурой ResNet-10 [5] в качестве основы. Его обучили на множестве изображений лиц из Интернета.

Рассмотрев данные алгоритмы, можно привести их ключевые преимущества и недостатки в таблице 1.

Прежде чем выбрать алгоритм для реализации проанализируем результаты тестирования приведенные в статье распознавания групповых эмоций [5] нужно определиться с требованиями, поставленными перед данным исследованием отходя от цели исследования. Ниже приведены результаты тестирования вышеуказанных алгоритмов на 300 случайно выбранных изображениях из обещающей выборки (100 изображений для каждого класса эмоций) (Рисунок 2):

Таблица 1 – Характеристика алгоритмов распознавания лиц.

<b>Название алгоритма</b>	<b>Преимущество</b>	<b>Недостаток</b>
HoG гистограммы ориентированных градиентов	1. Самый быстрый метод обнаружения лица. 2. Работает с изображениями с небольшими окклюзиями.	1. Не распознает лица размером менее 80x80 пикселей.

		2. Не работает с несовмещенными изображениями лиц и изображениями с большими окклюзиями.
Haar Cascade Face Detector	<ol style="list-style-type: none"> <li>1. Обнаружение почти в реальном времени.</li> <li>2. Обнаруживает лица в разных масштабах.</li> </ol>	<ol style="list-style-type: none"> <li>1. Множество ложных срабатываний.</li> <li>2. Не работает с невыровненными изображениями и изображениями с окклюзиями.</li> </ol>
MTCNN многозадачная каскадная сверточная нейронная сеть	<ol style="list-style-type: none"> <li>1. Работает в реальном времени на GPU.</li> <li>2. Очень точное обнаружение лица.</li> <li>3. Обнаруживает лица в разных масштабах.</li> <li>4. Работает для разной ориентации лица.</li> <li>5. Работает с изображениями с окклюзиями.</li> </ol>	<ol style="list-style-type: none"> <li>1. Медленная загрузка процессора.</li> </ol>
DNN - Deep Neural Network	<ol style="list-style-type: none"> <li>1. Очень точное распознавание лиц.</li> <li>2. Работает в режиме реального времени.</li> <li>3. Обнаруживает лица в разных масштабах.</li> <li>4. Работает как с невыровненными изображениями, так и с изображениями с окклюзиями</li> </ol>	<ol style="list-style-type: none"> <li>1. Игнорирует множество лиц, т.е. обнаруживает меньшее количество лиц по сравнению с другими алгоритмами.</li> </ol>
AdaBoost	<ol style="list-style-type: none"> <li>1. хорошая обобщающая способность.</li> <li>2. простота реализации;</li> <li>3. собственные накладные расходы бустинга невелики.</li> <li>4. возможность идентифицировать объекты, являющиеся шумовыми выбросами.</li> </ol>	<ol style="list-style-type: none"> <li>1. Бывает переобучение при наличии значительного уровня шума в данных.</li> <li>2. AdaBoost требует достаточно длинных обучающих выборок.</li> <li>3. Бывает построение неоптимального набора базовых алгоритмов.</li> <li>4. Бустинг может приводить к построению громоздких композиций,</li> </ol>

		состоящих из сотен алгоритмов.
--	--	--------------------------------

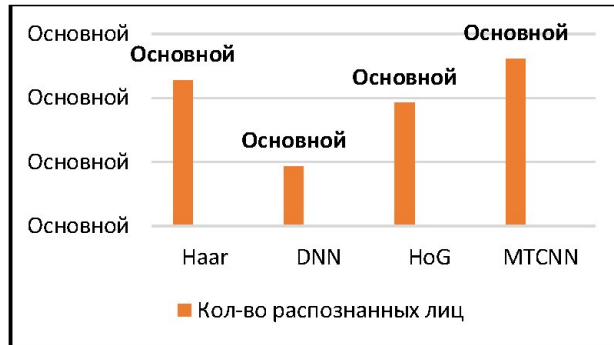


Рисунок 2 – Количество лиц, обнаруженных каждым методом на наборе 300 изображений из обучающего набора

В дальнейшем предстоит изучить некоторые характерные особенности изображения с методами распознавания, которые можно использовать с алгоритмами машинного обучения. Ниже приведены эти характерные особенности, которые можно использовать для обучения машинного обучения:

Система кодирования движений для лица (Facial Action Coding System FACS) относится к набору движений лицевых мышц, которые соответствуют отображаемой эмоции. Первоначально созданный Карлом-Германом Хьортсьё с 23 модулями движения лица в 1970 году, он был впоследствии развит Полом Экманом и Уоллесом Фризенем в 1978 году, и была пересмотрена Экманом, Фризенем и Хагером в 2002 году.

Кодировщики FACS описывают каждое наблюдаемое изменение движения лица на основе единиц действия (Action Unit(AU)). Они указывают, какие AU переместились, чтобы вызвать изменения, наблюдаемые на лице. Это делает кодирование FACS вполне объективным.

При подсчете единиц необходимо будет применять замедленное движение и покадровый просмотр для определения возникающих AU, всегда чередуя с просмотром в реальном времени. Таким образом, кодирование FACS требует очень много времени. Например, улыбающееся лицо можно определить с точки зрения единиц действия как движение мышцы AU6 (это устройство для поднятия щек) и мышцы AU12 (это устройство для снятия углов губ) приводит к счастливому лицу (Таблица 2).

Таблица 2 – Единицы действия, соответствующие различным движениям лица

AU6		Поднятие щек	AU12		Съемник уголков губ
-----	---	--------------	------	--	---------------------

Ориентиры (Landmarks) на лице очень важны и могут использоваться для обнаружения и распознавания лиц. Те же ориентиры можно использовать

и в выражениях. В библиотеке Dlib есть детектор 68 ориентиров на лице, который определяет положение 68 ориентиров на лице.

Используя библиотеку Dlib, можно извлечь координаты (x, y) каждой из точек лица. Эти 68 точек можно разделить на определенные области, такие как левый глаз, правый глаз, левая бровь, правая бровь, рот, нос и челюсть.

Еще один метод дескриптор функции описывает область вокруг ключевой точки, которые определяют углы и края на изображениях в библиотеке OpenCV с помощью вышеупомянутого алгоритма детектора признаков, таких как детектор углов Харриса. Описание может быть любым, включая необработанные значения интенсивности пикселей или координаты окружающей области.

После тщательного изучения статей предыдущих исследований было выявлено, что не целесообразно использовать одного алгоритма и комбинация алгоритмов распознавания HoG и многозадачного каскадных сверточных сетей (MTCNN) является одним из методов для достижения максимальной производительности и наивысшей точности в задачи распознавания лиц на изображениях.

#### **Список литературы:**

1. Jiaqi Shi, Chaoran Liu, Carlos Toshinori Ishi and Hiroshi Ishiguro « Skeleton-based emotion recognition based on two-stream self-attention enhanced spatial-temporal graph convolutional network», 2021
2. Yoon, S.; Dey, S.; Lee, H.; Jung, K. Attentive modality hopping mechanism for speech emotion recognition., Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 3362–3366.
3. Kołakowska, A.; Landowska, A.; Szwoch, M.; Szwoch, W.; Wrobel, M.R. Emotion recognition and its applications. In Human-Computer Systems Interaction: Backgrounds and Applications 3; Springer: Cham, Switzerland, 2014; pp. 51–62.
4. Nitisha Raut «Facial Emotion Recognition Using Machine Learning», San Josй State University, May 2018
5. Samanyou Garg «Group emotion recognition using machine learning. Third year project report» School of computer science the university of manchester United Kingdom, April 2019.