

# РАСПОЗНОВАНИЕ И КЛАССИФИКАЦИЯ ЛИЦЕВЫХ ЭМОЦИЙ НА ОСНОВЕ ВИЗУАЛЬНОЙ ИНФОРМАЦИИ НА ВИДЕОПОТОКЕ

В.А. Коровкин

Томский политехнический университет

E-mail: alcasar@tpu.ru

## Введение

Сегодня активно развиваются различные способами человеко-машинного взаимодействия в реальном времени – голосовые помощники (например, Siri, Cortana), многочисленные боты, автоматизированные системы. Однако сложность заключается в том, что вербальными методами люди передают менее трети информации в процессе взаимодействия между собой. Согласно различным исследованиям в процессе коммуникации на бессознательном уровне активно используются невербальные каналы информации (тон голоса, поза тела, положение рук и т.д.) [1,2] Среди всех невербальных компонент, участвующих при межличностном общении, выражения лица человека является одним из самых информативных каналов. Так как большинство людей не контролируют лицевые эмоции и микровыражения лица, то во время разговора можно точно охарактеризовать эмоциональное состояние как участника, так и его отношение к предмету обсуждения и остальным участникам. Согласно исследованиям П. Экмана, которые были неоднократно подтверждены на практике существуют набор базисные эмоции, которые не зависят от культурного и географического контекстов, то есть выражаются одинаково у всех людей (например, грусть, радость, ненависть, отвращение и т.д.). Поэтому остро стоит вопрос о реализации систем, работающих в режиме реального времени, которые способны находить, классифицировать и анализировать эмоции людей по выражению лица. Данная задача относится к группе задач распознавания (выделения и классификации) образов на видеопотоке. Сегодня результаты исследований под данной теме только начинают и применяются на практике не только непосредственно в робототехнике, но и в таких сферах как компьютерная графика (процедурное динамическое создание различных графических материалов), автоматический анализ людей и создания психологического профиля (например, автоматическое определения усталости водителя), анимация, психология, моделирование поведение людей, безопасность и т.д.

## Задача

Для успешного решения задачи классификации эмоций на видеопотоке необходимо было выполнить ряд отдельных подзадач:

1. предобработка кадра;
2. детектирования лица на изображении (кадре);

3. извлечение признака на кадре (или на детектированном лице), на основе которого будет производиться классификация;
4. нормализация и создания векторов признаков;
5. непосредственно классификация полученной эмоции и анализирование результата.

Стоит также учитывать, что в кадре в единый момент времени может находиться несколько человек. Для классификации использовались 7 классов эмоций и нейтральное выражение лица.

## Описание выборки данных

Для обучения и проверки были использованы несколько наборов данных: FER2013 и Extended Cohn-Kanade (СК+).

Набор данных FER содержит более 35 000 изображений лица для выделенных 7 классов эмоций. Стоит отметить, что при подготовке данных был замечен переизбыток данных в категориях счастье и отвращение.

В наборе Extended Cohn-Kanade (СК+) также можно выделить 7 необходимых классов по аналогии с FER. Из данной базы также было взято более 30 000 изображений.

Для решения проблемы увеличения базы для обучения и тренировки в классах, у которых было меньше всего изображений было принято выполнить зеркальное отображение каждого такого изображения.

Наборы данных были разделены на две части – обучающую и тестовую в соотношении  $\approx 85\%$  и  $\approx 15\%$ .

Перед непосредственно обучением выборка фотографий была нормализована. Данные изображения должны обладать следующими свойствами:

- лицо расположено без сильных поворотов и наклонов (поворот составляет менее 10 %);
- фотография переведена в градации серого;
- лицо занимает как минимум около 90% изображения.



Рис. 1. Примеры изображений из датасета

### Описание подхода

Для детектирования лица в кадре можно воспользоваться двумя методами: методом Виолы-Джонса и Eigenfaces. В данной работе был использован алгоритм Виолы-Джонса. Обучение алгоритма осуществлялось на основе бустинга, при котором классификаторы (признаки Хаара) комбинируются таким образом, чтобы получить более сложный классификатор. Для обучения использовались наборы из общедоступных баз данных. Eigenfaces (собственные лица) – это название совокупности собственных векторов, используемых в задаче компьютерного зрения распознавания человеческих лиц. Данный подход использует метод главных компонент. Данный алгоритм больше подходит для идентификации человека по фотографии (то есть нахождения соответствия с эталоном).

Существует несколько способов определения эмоций. Один из методов определения эмоций по лицу человека основан на классификации ключевых точек (facial landmarks). Координаты данных точек были получены через алгоритм Active Appearance Model. Для решения было выделено 78 ключевых точек. Один из способов определения эмоций по изображению лица основан на классификации ключевых точек (facial landmarks). Координаты данных точек были получены через алгоритм Active Appearance Model. Для решения было выделено 78 ключевых точек. Общая схема работы алгоритма представлена на рисунке 2.

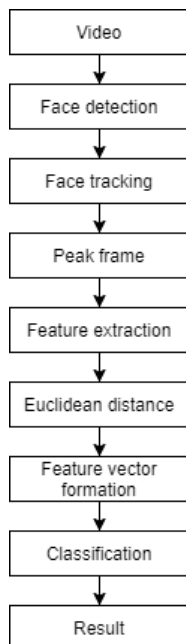


Рис. 2. Блок-схема работы алгоритма

Для реализации непосредственно классификации были использованы последовательно несколько типов нейронных сетей: сверточные сети (CNN) и рекуррентные сети (LSTM, RNN). Сверточная сеть впервые очередь используется для анализа

признаков через операции свертки и непосредственно классификации эмоции. Рекуррентная сеть позволяет оптимизировать работу последовательности изображений (видеопотока). Таким образом, возможно анализировать только кадры с «крупными» изменениями и синхронизировать кадры, которые используются непосредственно в сверточной нейронной сети.

Проект разрабатывался на языке python с использованием библиотеки Keras с Tensorflow.

### Заключение

В данной работе были рассмотрены основные методы и способы работы реализации алгоритма классификации изображения на видеопотоке. Стоит отметить, что использование нескольких способов и комбинация нескольких признаков сильно повышают точность работы группы сетей.

В результате работы получили среднюю точность для семи эмоций в 70% и около 94% для трех. Исходя из результатов можно сделать, что происходит переобучения сети на большом количестве данных, реализована грубая модель взаимодействий между слоями нейронных сетей (то есть используется и передается избыточное число параметров в сетях).

В дальнейшем планируется оптимизировать работы сетей, разделить классификацию эмоций по нескольким признакам.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-08-00977 А и в рамках Программы повышения конкурентоспособности ТПУ.

### Список использованных источников:

1. Коровкин В.А. Современные методы распознавания эмоций / Коровкин В.А. - Молодежь и современные информационные технологии, 2018. – 76 -78 с.
2. Коровкин В.А. Применение нейронных сетей для классификации эмоций на изображениях / Коровкин В.А. – Тезисы xx всероссийской конференции молодых учёных по математическому моделированию и информационным технологиям, 2019. – 62 – 63с.
3. Экман П. Психология эмоций / П Экман. – Спб.: Питер, 2010. – 336 с.
4. Chu, W.S.; Torre, F.D.; Cohn, J.F. Learning spatial and temporal cues for multi-label facial action unit detection. In Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 30 May–3 June 2017; pp. 1–8
5. Jain, D.K.; Zhang, Z.; Huang, K. Multi angle optimal pattern-based deep learning for automatic facial expression recognition. Pattern Recognit. Lett. 2017, 1, 1–9