

УДК 004.92, 004.93

Анализ методов распознавания эмоций по фотоизображениям

А. П. Семёнова, М. А. Белинская, В. Н. Павлыш
Донецкий национальный технический университет
nastena-semenova19@rambler.ru

Аннотация

В работе рассмотрен актуальный на сегодняшний день метод обнаружения лица на изображении. Проанализированы различные подходы для классификации эмоций, как классический, предусматривающий использование ключевых точек, так и с применением глубокого обучения. Определены модели категоризации эмоций: дискретная, многомерная и гибридная. Сделан вывод, что для получения более точных результатов необходимо применять комбинированные методы с использованием различных входных данных.

Введение

Распознавание эмоций является одной из самых обсуждаемых тем в области машинного обучения. Задача распознавания эмоций является трудно формализуемой, и потому при ее постановке часто возникает ряд проблем, главными из которых являются отсутствие стандарта мимических картин и субъективность мнения эксперта, поскольку человек не может в полной мере воспринимать все микродвижения собеседника. Однако, опираясь на исследования специалистов, можно выделить несколько универсальных эмоций, мимические проявления которых не зависят от расы и культуры человека [1-8].

Распознавание эмоций – это задача, решение которой может служить для совершенно разных целей. Она может применяться, как для маркетинга, систем видео аналитики, для создания мультимедиа приложений и игр, так и для более сложных задач, таких как анализ психического и эмоционального состояния людей, мониторинга учащихся и т.д. [9].

Поиск лица на изображении

Первым шагом к распознаванию эмоций на лице является обнаружение самого лица на фотоизображении. Рассмотрим наиболее популярный метод обнаружения лиц – метод Виолы-Джонса. Этот метод обеспечивает высокую точность и достаточно быструю скорость нахождения объектов на изображении.

Метод Виолы-Джонса (Viola-Jones object detection) – алгоритм обнаружения объектов на изображении в реальном времени, разработанный в 2001 году П. Виолой и М. Джонсоном [10,11]. Несмотря на то, что главной задачей этого метода является нахождение лиц, его используют и для распознавания других объектов. Существует множество реализаций, в том числе в составе библиотеки компьютерного

зрения OpenCV (например, функция `cvHaarDetectObjects()`).

Метод Виолы-Джонса использует принцип сканирующего окна, который состоит из пяти основных этапов:

1) поступление исходного изображения размерностью $N \times M$ пикселей на вход, при этом все пиксели имеют значение от 0 до 255 для каждого цветового канала (канал одного цвета для монохромного и трех цветов для цветного изображения);

2) обработка изображения (применение различных фильтров, масштабирование, удаление шумов и т.д.);

3) сканирование изображения при помощи скользящего окна;

4) проход скользящего окна по каждому пикселю изображения;

5) применение классификатора на основе метода бустинга (усиления слабых классификаторов).

Процесс обнаружения лица методом Виолы-Джонса имеет ряд особенностей:

- изображения применяют в интегральном виде для быстроты расчетов;

- проводится анализ признаков Хаара для поиска нужных объектов на изображениях;

- применяется метод усиления слабых классификаторов на определенной части изображения для выбора наиболее подходящих признаков при поиске объекта;

- применяются простые бинарные классификаторы для принятия решения «true» или «false»;

- применяются каскады признаков для быстрого отбрасывания окон без искомого объекта.

Метод позволяет выполнить быстрое вычисление суммарной яркости произвольного прямоугольника с постоянным временем, независимо от размеров этого прямоугольника благодаря обработке изображения в интегральной форме. Представление

изображения происходит в матричном виде (размер матрицы совпадает с размером изображения). В каждом элементе матрицы хранится сумма интенсивностей всех пикселей, которые находятся левее и выше данного элемента.

Применяемые алгоритмом признаки используют суммирование пикселей из прямоугольных областей и содержат более одной прямоугольной области (рис. 1).

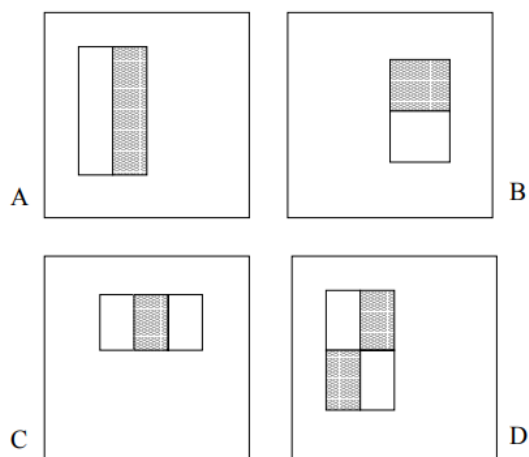


Рисунок 1 – Признаки алгоритма Виолы-Джонса

На рисунке 1 приведены четыре различных варианта признаков. Для вычисления значения каждого признака нужно из суммы пикселей белых прямоугольников вычесть сумму пикселей серых областей.

Прямоугольные признаки примитивнее, чем поворачивающийся фильтр, и результат их поиска грубее. Но их преимущество состоит в том, что благодаря хранению изображения в интегральной форме, проверка прямоугольного признака на конкретной позиции выполняется за константное время.

Каждая прямоугольная область используемых признаков всегда смежна с другой прямоугольной областью, поэтому расчет признака с двумя прямоугольными областями состоит из шести обращений в интегральный массив, для признака с тремя прямоугольными областями – из восьми, а с четырьмя – из девяти.

Преимуществами данного алгоритма являются:

- эффективный набор признаков;
- быстрое вычисление значений признаков;
- масштабирование признаков вместо масштабирования изображения.

К недостаткам алгоритма относятся:

- чувствительность к углу падения света;
- невозможность распознать требуемый

объект при его повороте более чем на 45 градусов;

– значительные временные затраты на обучение классификаторов.

Категоризация эмоций

Следующим этапом распознавания эмоций является их категоризация. Сегодня существуют три наиболее популярных модели категоризации эмоций: дискретная, гибридная и многомерная [1,5,12].

Дискретная модель основана на классификации эмоций на основе естественного языка. Каждая эмоция связана с семантическим параметром, то есть обладает конкретным значением или набором значений. Теория базовых (универсальных) эмоций является ярким примером данного подхода. Таким образом, например, ностальгия – это сочетание таких базовых эмоций как радость и печаль.

Различные исследователи выделяют разное количество и типы базовых эмоций. П. Экман выделяет 6 первичных эмоций (гнев, страх, отвращение, удивление, печаль и радость), психозволюционная теория Р. Плутчика рассматривает 8 базисных эмоций (одобрение, печаль, ожидание, отвращение, радость, страх, гнев, удивление), а Дж. Грей – 3 основные эмоции. По мнению, Мауэра следует говорить, только о 2 первичных эмоциях: боль и удовольствие.

На рисунке 2 изображена модель эмоций Р. Плутчика.

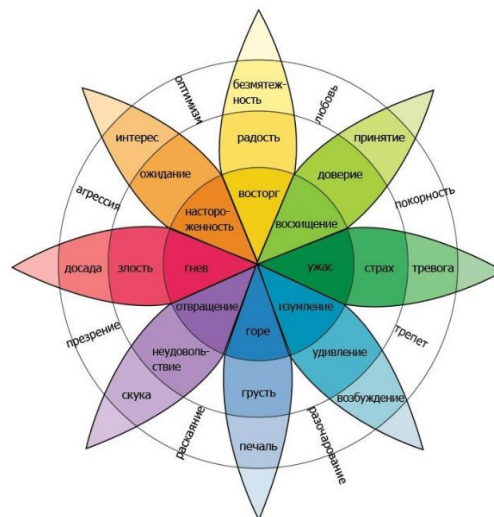


Рисунок 2 – Колесо эмоций Р. Плутчика

Многомерная модель представляет эмоции в координатном многомерном пространстве. Так как пространство является неразрывным и непрерывным, то эмоции различаются по одному или нескольким параметрам (при этом они могут иметь одинаковую основу).



Рисунок 3 – Двумерная модель эмоций Дж. Рассела

Ярким примером описываемого подхода является модель Дж. Рассела (см. рис. 3). В ней водится двумерный базис, в котором каждая эмоция характеризуется знаком (valence) и интенсивностью (arousal). Данная модель в контексте решение задачи классификации является одной из самых востребованных.

Гибридные модели объединяют как дискретные, так и многомерные подходы. Отличным примером, иллюстрирующим данную модель, являются «Песочные часы эмоций», которые были предложены Камбрией, Ливингстоном, Хуссейном. Каждое измерение характеризуется 6 уровнями силы, с которой выражены эмоции. Данные уровни обозначаются набором из 24 эмоций. Поэтому совершенно любая эмоция может рассматриваться, как фиксированное состояние, так и часть пространства, связанная с другими эмоциями нелинейными отношениями.

Определение ключевых точек

На сегодняшний день, один из популярных способов определения эмоций основан на классификации по ключевым точкам на изображении лица [13]. Существует множество алгоритмов поиска ключевых точек на изображении. Первые методы поиска основаны на подборе параметров деформации для лицевой модели. Наиболее известные классы алгоритмов включают Constrained Local Model (CLM), Active Appearance Model (AAM), Active Shape Model (ASM) [14,15]. С помощью деформируемой модели лица вычисляются координаты каждой ключевой точки. Часто такие алгоритмы используют в качестве основы статистические методы. Они достаточно хороши на наборах данных в

контролируемых условиях (правильное освещение, наклон головы и т.д.), но плохо справляются с обработкой зашумленных изображений. Лучший результат поиска ключевых точек лица на наборах данных с большой вариацией поз и частичным перекрытием показывают регрессионные нейросетевые методы.

Чаще всего на изображении определяют от 5 до 68 точек, которые привязывают к положению бровей, глаз, носа, рта и челюсти. Данная привязка позволяет захватить мимику объекта. На рисунке 4 представлена модель с использованием 68 и 10 ключевых точек.

Полученные координаты точек нормируются и непосредственно передаются в классификатор. В роли классификатора может выступать метод опорных векторов (SVM) или, например, Random Forest.

Для распознавания эмоции необходимо иметь представление о точном положении ключевых точек лица. Для понимания процесса их поиска рассмотрим несколько ключевых понятий.

Двигательная единица (action units) – величина, которая принимает значения из диапазона от 0 до 1, и определяет насколько далеко находится конкретная ключевая точка от линии глаз или нормали к этой прямой. Если значение двигательной единицы принимает значение больше 0, то ее называют активированной.

Нейтральное расстояние – величина, определяющая среднее расстояние от конкретной ключевой точки до нужной прямой при нейтральном выражении лица, т.е. когда коэффициенты проявления каждой конкретной эмоции принимают значения не более 0,26.

Максимальное расстояние – расстояние от прямой до ключевой точки, относящейся к двигательной единице связанной с проявлением той или иной эмоции. Стрелки на рис. 5

показывают направления для изменения положения ключевых точек (в данном примере таких точек 10: 2 для зрачков, по 4 для бровей и губ) для активации двигательных единиц.

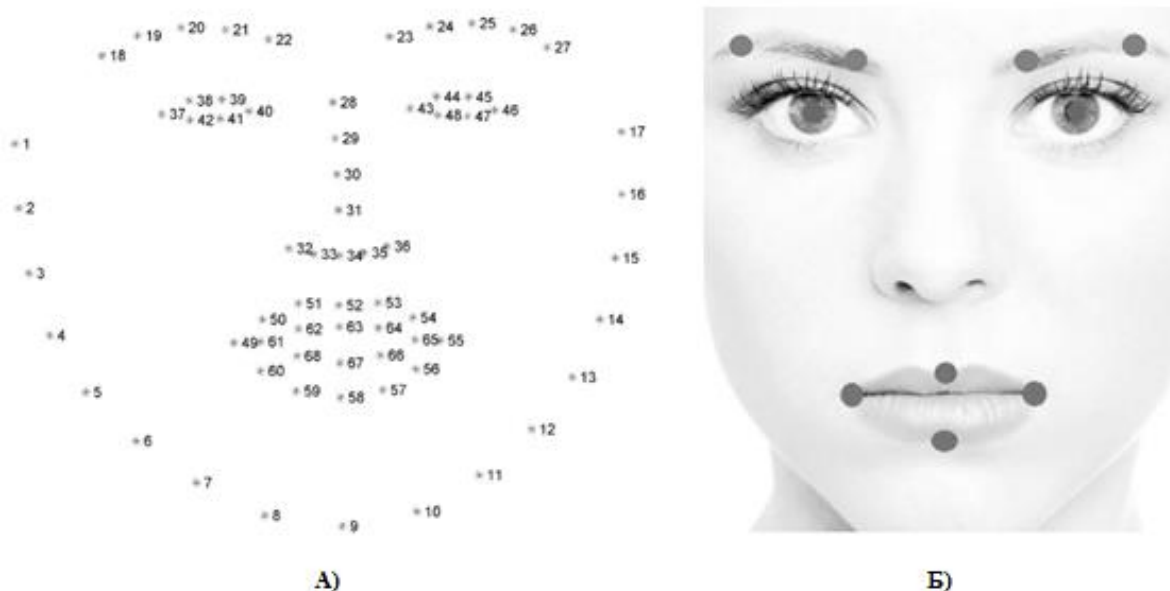


Рисунок 4 – Ключевые точки лица (А) – 68 точек, Б) – 10 точек)

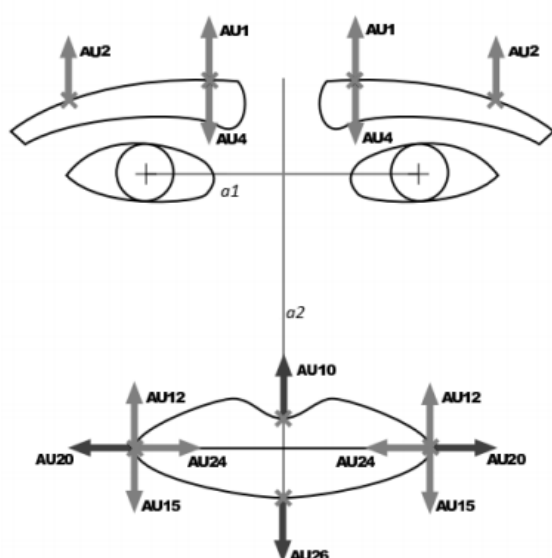


Рисунок 5 – Двигательные единицы

Для каждой точки определяется область интереса на анализируемом фотоизображении. Чтобы уменьшить зону поиска и снизить время на их нахождение, эти области определяют с помощью различных классификаторов для обнаружения зон лица, рта, глаз и т.д.

Для поиска ключевых точек в зоне бровей можно использовать адаптивный контрастный фильтр, задав ему минимальное и

максимальное значения серого цвета, чтобы обнаружить границы бровей и представить их в виде набора соединённых точек или полигона. Для поиска зрачков можно использовать классификатор, позволяющий определить точное положение глаз и зрачков на фотоизображении. Для определения положения остальных четырёх точек нужно, используя классификатор, разбить зону поиска на 4 части, а после применить адаптивный контрастный фильтр для нахождения граничных точек.

Для классификации эмоции по ключевым точкам используется система кодирования лицевых движений (Facial Action Coding System), разработанная П. Экманом и У. Фризеном [16,17].

После определения всех ключевых точек вычисляются значения двигательных единиц. Для этого задаются нейтральное и максимальное расстояния от ключевых точек до прямой, проходящей через координаты зрачков или нормали к этой прямой. Средние значения этих расстояний можно получить с помощью анализа нескольких фотоизображений.

Далее на анализируемом фотоизображении необходимо рассчитать расстояния от ключевых точек до нужных прямых и сравнить с нейтральным и максимальным расстояниями.

Так как, размеры фотоизображений могут отличаться, то за 1 нужно взять расстояние между зрачками. Если полученное расстояние

больше или равно максимальному, то значение двигательной единицы становится равным 1, если оно меньше нейтрального, то – 0, а если находится между нейтральным и максимальным расстоянием, то значение вычисляется с помощью линейной интерполяции.

Для определения конкретных эмоций необходимо сложить полученные значения двигательных единиц, отвечающих за проявление конкретной эмоции и разделить на общее количество взятых единиц. Таким образом, получится число в диапазоне от 0 до 1, показывающее степень выраженности той или иной эмоции.

Классификация эмоций с применением глубинного обучения

При анализе визуальных данных лучшим выбором является использование глубоких сверточных нейронных сетей [18-20]. Мгновенные снимки не достаточно точно отражают текущую эмоцию, поэтому для определения эмоции необходимо использовать последовательность кадров. В качестве нейросетевого классификатора можно использовать сеть для распознавания лиц.

Решение задачи анализа последовательности кадров можно решить двумя способами: использование рекуррентной сети (LSTM) или использование архитектуры 3D-CNN. В первом варианте входные данные необходимо подать на сверточную сеть. Высокоуровневые признаки, полученные от CNN, которая классифицирует каждый отдельный кадр, затем передаются в рекуррентную сеть. В ней происходит обработка временной составляющей. Во втором случае непосредственная подача последовательности кадров с некоторым шагом выполняется прямо на входе сверточной нейронной сети 3D-CNN. Данная сеть использует свертки с тремя степенями свободы, которые преобразует четырехмерный вход в трехмерные карты признаков.

Заключение

В работе рассмотрен основной и актуальный на сегодняшний день метод обнаружения лица на изображении – метод Виолы-Джонса. Данный метод имеет ряд таких достоинств как:

- быстрое вычисление значений признаков;
- эффективный набор признаков;
- масштабирование признаков вместо изображения.

Однако он не универсален и имеет свои требования к исходному изображению.

Проанализированы различные подходы

для классификации эмоций, как классический с использованием ключевых точек, так и с применением глубинного обучения. Наиболее лучшие результаты показывают методы с применением сетей глубинного обучения (CNN). Существуют также иные методы, с использованием аудио (речь) и видео (лицо и мимика) материалов. Для получения более точных результатов необходимо применять комбинированные методы с использованием различных входных данных.

Литература

1. Леонтьев, В. О. Классификация эмоций / В. О. Леонтьев. – Одесса: Инновационно-ипотечный центр, 2002. – 84 с.
2. Ильин, Е. П. Эмоции и чувства / Е. П. Ильин. – СПб.: Питер, 2001. – 752 с.
3. Джеймс, У. Психология / У. Джеймс. – М.: Педагогика, 1991. – 368 с.
4. Ланге, Г. Душевные движения / Г. Ланге. – СПб.: Питер, 1996. – 180 с.
5. Миненко, А. С. Формальная модель эмоций / А. С. Миненко, А. П. Семенова // Проблемы искусственного интеллекта, 2018. – №3(10). – С. 84-93.
6. Миненко, А. С. Моделирование и информационные технологии при распознавании лица человека по его мимическим изображениям // Проблемы искусственного интеллекта, 2016. – № 2(3). – С. 48-54.
7. Семенова, А. П. Математическая модель эмоций // Международная научно-техническая конференция молодых ученых БГТУ им. В.Г. Шухова. Материалы национальной конференции с международным участием. – Белгород, 2019. – С. 4584-4587.
8. Семёнова, А. П. Анализ мимических выражений для задачи распознавания эмоций / А. П. Семёнова, В. Н. Павлыш // Проблемы искусственного интеллекта, 2020. – № 4(19). – С. 69-79.
9. Семенова, А. П. Область применения алгоритма распознавания эмоций в информационных технологиях / А. П. Семенова, А. С. Миненко, Т. В. Ванжа // Цифровой регион: опыт, компетенции, проекты: сборник статей Международной научно-практической конференции (г. Брянск, 30 ноября 2018 г.). – Брянск: Брян. гос. инженерно-технол. ун-т., 2018. – С. 443-446.
10. Yang, Ming-Hsuan. Detecting Faces in Images: a Survey / Ming-Hsuan Yang // IEEE Trans. Pattern Analysis and Machine Intelligence – 2002. – № 11. – P. 34–58.
11. Viola, P. Robust real-time face detection / P. Viola. – International Journal of Computer Vision, 2005. –V. 58. – № 2. – P 137–155.

12. Экман, П. Психология эмоций. Я знаю, что ты чувствуешь. / П. Экман – СПб. : Питер, 2010. – 334 с.
13. Семёнова, А. П. Поиск ключевых точек лица для задачи распознавания эмоций / А.П. Семёнова, В.Н. Павлыш // Информатика и кибернетика, 2021. – № 1-2 (23-24). – С. 59-64
14. Wang, N. Facial feature point detection: A comprehensive survey / N. Wang [et al.] // Neuro computing, 2018. – №275. – P. 50-65.
15. Wu, Y. Facial Landmark Detection: A Literature Survey / Y. Wu, Q. Ji // International Journal of Computer Vision, 2018. – №.127. – P. 115-142.
16. Friesen, W. EMFACS-7: Emotional Facial Action Coding System. Unpublished manual / W. Friesen, P. Ekman. – California: University of California, 1983.
17. Ekman P. Facial Action Coding System: A Technique for the Measurement of Facial Movement / P. Ekman, W. Friesen. – Palo Alto: Consulting Psychologists Press, 1978. – 117 p
18. Семёнова, А. П. Анализ методов распознавания выражений лица / А. П. Семёнова, А. С. Миненко // Искусственный интеллект: теоретические аспекты, практическое применение: материалы Донецкого международного научного круглого стола. – Донецк: ГУ ИПИИ, 2020. – С.192-197.
19. Subject independent facial expression recognition with robust face detection using a convolutional neural network [Текст] / М. Masakazu, K. Mori, Y. Mitari, Y. Kaneda // Neural Networks, 2003. – № 5 (16). – P. 555-559.
20. Facial expression recognition via a boosted deep belief network. [Текст] / P. Liu, S. Ha, Z. Meng, Y. Tong // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 24–27 June 2014. – P. 1805–1812.

Семёнова А. П., Белинская М. А., Павлыш В. Н. Анализ методов распознавания эмоций по фотоизображениям. В работе рассмотрен актуальный на сегодняшний день метод обнаружения лица на изображении. Проанализированы различные подходы для классификации эмоций, как классический, предусматривающий использование ключевых точек, так и с применением глубинного обучения. Определены модели категоризации эмоций: дискретная, многомерная и гибридная. Сделан вывод, что для получения более точных результатов необходимо применять комбинированные методы с использованием различных входных данных.

Ключевые слова: метод Виолы-Джонса, базовые эмоции, ключевые точки лица, двигательные единицы, глубинное обучение.

Semenova A.P., Belinskaya M.A., Pavlysh V.N. Analysis of emotion recognition methods based on photo images. The paper considers the current method of detecting a face in an image. Various approaches for the classification of emotions are analyzed, both classical, involving the use of key points, and with the use of deep learning. Emotion categorization models are defined: discrete, multidimensional and hybrid. It is concluded that in order to obtain more accurate results, it is necessary to apply combined methods using different input data.

Key words: Viola-Jones object detection, basic emotions, key points of the face, action units, deep learning.

Статья поступила в редакцию 25.05.2022
Рекомендуется к публикации профессором Мальчевой Р. В.