ОБЗОР МЕТОДОВ КЛАССИФИКАЦИИ ЭМОЦИЙ ПО ВИДЕОИЗОБРАЖЕНИЮ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Жабинский А.В.

Одинец Д.Н. − к. т. н., доцент

Построение полноценных систем взаимодействия между человеком и компьютером подразумевает использование не только вербальной, но и невербальной информации. Одним из наиболее значимых источников невербальной информации при общении между людьми является выражение лица собеседника. В связи с этим получили распространение методы автоматического определения эмоций по видеоизображению лица.

Все такие методы можно условно разделить на две группы — статические и динамические. Статические методы работают с отдельными кадрами и направлены прежде всего на определение выражения лица в конкретный момент. В отличие от них, динамические методы рассматривают сразу последовательность кадров и делают упор на изменение характеристик лица во времени. Достоинством динамических методов является то, что некоторые эмоции определены только через изменения. Например, испуг характеризуется как кратковременное поднятие бровей и расширение глаз. В то же время, удержание такого выражения лица более 1 секунды скорее указывает на удивление [1].

Один из первых методов, основанных на изменениях во времени, использовал **оптические потоки** для выявления движения мышц лица [3]. Данный метод позволяет на основе яркости пикселей получить векторы движения, а также их величину (ускорение). Затем полученные данные сравнивались с образцами для каждой исследуемой эмоции, и выбиралась та из них, которая давала наибольшую вероятность совпадения. Сам по себе данный метод не получил широкого распространения из-за ряда ограничений, таких как неустойчивость к поворотам головы, необходимость предварительно локализовать положение лица на изображении, а также невозможность отслеживать статические выражения. Тем не менее, использованный подход стал основой для развития многих других методов и на данный момент часто цитируется в соответсвующих научных работах.

Другой динамический метод использует **модель соединённых вибраций** [4]. Суть метода состоит в наложении на интересующие регионы (такие как область глаз и рта) специальной сетки (мембраны). Образованные сеткой участки «закрепляются» за соответсвующими участками лица, а на основе последовательности кадров для каждого участка выводится модель вибраций, описывающих возможные трансформации. Для улучшения модели используется фильтр Кальмана, позволяющий уменьшить шумы и восстановить неполные результаты. Сама по себе классификация эмоций, как и в случае с предыдущим методом, строится на основе сопастовления с образцом.

Существуют и более сложные методы распознавания эмоций на основе модели вибраций. Так, например, возможно использовать **скрытые марковские модели** [5] для введения дополнительных ограничений, описывающих вероятности переходов между выражениями лица. Дополнительным плюсом является возможность автоматической сегментации потока кадров для выделения периодов, соответсвующих определённым эмоциям.

Ограничением всех динамических моделей является невозможность работать со статическими изображениями, что сильно уменьшает количество вариантов использования. Например, если для исследования доступно только одно изображение или короткая последовательность кадров, в течение которых выражение лица не меняется (радость, гнев и большинство других эмоций характеризуется относительной устойчивостью во времени), для динамических методов будет просто недостаточно признаков для классификации. Кроме того, описанные выше методы и модели плохо поддаются адаптации для использования в статическом контексте.

С другой стороны, существуют методы, изначально опирающиеся на признаки, не требующие последовательного набора кадров, и поэтому способные работать со статическими изображениями. Как правило, используются признаки изображения из временной области, хотя существуют и работы, использующие признаки из частнотной области [6].

Наиболее простой способ распознавания эмоций по статичному кадру основан на методе Eigenfaces и представляет из себя просто классификатор, входами которого являются значения пикселей внутри интересующией области. Очевидно, что этот метод не захватывает никаких специфичных для выражения эмоций (да и вообще лица) признаков и в принципе не способен дать хорошие результаты, поэтому он никогда не был реализован на практике. Немного более сложным является подход, основанный на хааровских признаках: на интересующие области лица накладываются фильтры, позволяющие подсчитать разницу между освещённостью участков. Данный подход также не получил широкого распространения, однако часто используется как вспомогательный метод локализации лица и его элементов.

Из предыдущих примеров очевидно, что для достижения высокой точности классификации необходимо использовать признаки, которые бы отражали изменения на лице человека. Считается, что люди определяют выражение лица собеседника по форме и положению его элементов (в т.ч. глаз, бровей, носа и рта), а также, хотя и в меньшей степени, по цвету и текстуре кожи.

Форму и положение элементов лица можно задать, например, через соответсвующие контуры.

Именно такой подход используется в методе, основанном на применении **кривых Безье** [7]. Данный метод работает следующим образом. Вначале изображение сегментируется по цвету кожи, что позволяет определить открытые участки тела. Далее среди выделенных участков локализуется лицо, а также заданные элементы (глаза, рот), которые затем аппроксимируются кривыми Безье. Полученные коэффициенты кривых в последствии сравниваются с образцами из базы данных, после чего выбирается эмоция с наибольшей степенью совпадения.

Более простым и, в то же время, более надёжным способом определения элементов лица является использование ориентиров (ключевых точек). Именно такой способ используется в **моделях активного образа** (Active Appearance Models) [8]. Суть метода состоит в следующем. Вначале подготавливается обучающая выборка, состоящая из набора размеченных изображений, т. е. для каждого изображения подготавливается список ориентиров фиксированного размера. Каждый ориентир представляет из себя 2D или 3D точку, описывающую ключевую точку значимого контура (например, контур брови может быть задан 3-мя или более такими точками, контур глаза — 4-мя и т. д.). Обязательным является описание с помощью ориентиров внешнего контура лица.

Набор ключевых точек называется формой. Все полученные из обучающей выборки формы выравниваются при помощи прокрустова анализа: вначале формы сдигаются в начало координат, затем масштабируются и поворачиваются вокруг своей оси, чтобы соответсвовать некой базовой форме. Затем вычисляется средняя форма и процесс повторяется с использованием её в качестве базы.

После выравнивания формы «разворачиваются» в одномерный вектор. Например, если форма состояла из M точек в двумерном пространстве, то развёрнутая форма будет представлена вектором размера 2M (или, что то же самое, одной точкой в пространстве размерности 2M). Полученные вектора N форм собираются в матрицу размера N*2M, после чего с помощью метода главных компонент (Principal Component Analysis) уменьшается размерность пространства и выделяются наиболее важные (несущие наибольшее количество информации) оси координат. Наконец, задаётся модель формы:

$$S = S_0 + R_s p \quad (1)$$

где S — генерируемая форма, S_0 — средняя форма, R_S — матрица преобразования, полученная из первых k главных компонент и p — вектор параметров модели.

Аналогичным образом задаётся и модель текстуры: все пиксели, попавшие внутрь внешнего контура формы, организуются в виде многомерного вектора; к матрице, состоящей из таких векторов, применяется метод главных компонент, и задаётся сама модель текстуры:

$$T = T_0 + R_t p \quad (2)$$

Две эти модели вместе и задают модель активного образа.

При подгонке модели к неразмеченному изображению область в пределах внешнего контура триангулируется (как правило, методом Делоне), в результате чего получается своеобразная сетка. Затем каждый треугольник внутри этой сетки деформируется с помощью аффинного проеобразования и выравнивается к соответсвующему треугольнику модели. Такое преобразование даёт возможность вычислить «изображение ошибки» - разницу между текстурой модели и текстурой реального изображения. На основе величины этой ошибки вычисляются поправки к параметрам модели. Затем этот процесс повторяется до достижения стабильного минимума ошибки.

Модель активного образа позволяет достаточно точно определить положение ключевых точек на неразмеченном изображении, однако сама по себе она не покрывает область распознавания эмоций. Можно выделить несколько подходов к решению этой проблемы, например, использование обучения с учителем с использованием расстояний между ключевыми точками в качестве признаков или кластеризация/классификация на основе нейронных сетей и пр. Однако сама по себе модель активного образа подсказывает другой, статистический подход: для каждого ориентира можно создать модель распределения точки (Point Distribution Model), описывающую вероятность появления точки в том или ином месте при каждой из эмоций. Затем на основе теоремы Байеса можно создать простой классификатор, выбирающий наиболее вероятную эмоцию при заданном наборе положений ориентиров.

Следует отметить, что модели активного образа во многом напоминают модели соединённых вибраций, хотя и используют, во-первых, определённый набор ключевых точек вместо более общей мембраны, а во-вторых, набор несвязанных (в общем случае) изображений некоторого объекта. Тем не менее, AAM можно легко расширить также и для использования временной информации о субъекте исследования, что значительно увеличивает возможности развития модели.

Список использованных источников:

1. Экман П., Дарвин Ч. О выражении эмоций у человека и животных / П. Экман, Ч. Дарвин. - Питер, 2013.
2. Экман П. Психология лжи. Питер, 2009.
3. Mase K. An application of optical flow. Extraction of facial expression / K. Mase // NTT Human Interface Laboratories. - Tokyo, 1990.
4. Tao H. Connected vibrations: a model analysis approach for non-rigid motion tracking / H. Tao, T. S. Huang // Backman Institute.
5. Cohen I. Emotion recognition from facial expression using multilevel HMM / Ira Cohen, T. S. Huang // Backman Institute.
6. Bouzalmat A. Facial face recognition method using Fourier transform filters Gabor and R_LDA / A. Bouzalmat // International Conference of Intelligent Systems and Data Processing, 2011.
7. Khan M. I. Facial expression recognition for human-robot interface / M. I. Khan, Md. Al-Amid Bhuiyan // Chittagong University of Engineering and Technology. - Bangladesh, 2009.
8. Cootes T. F., Edwards G. J., Taylor C. J. Active appearance models / Cootes T. F., Edwards G. J., Taylor C. J. // In Proc. European Conf. on Computer Vision, volume 2, pages 484–498. Springer, 1998.