

УДК 519.254

МЕТОД РАСПОЗНАВАНИЯ ЭМОЦИЙ НА ОСНОВЕ МОДЕЛИ РАСПРЕДЕЛЕНИЯ КЛЮЧЕВЫХ РАССТОЯНИЙ

А.В. ЖАБИНСКИЙ

Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь

Поступила в редакцию 12 ноября 2013

Предложен метод распознавания эмоций человека по изображению его лица. В основе метода лежит разработанная статистическая модель эмоций, а также вероятностный алгоритм, основанный на наивном байесовском классификаторе. В качестве математической модели задачи применяется модель активного образа, которая основана на оценке расстояний между наборами ключевых точек.

Ключевые слова: распознавание эмоций, модели активного образа, модель распределения ключевых расстояний.

Введение

Выражение эмоций и понимание психологического состояния человека является одним из ключевых факторов взаимодействия между людьми. Подсчитано, что семантическая составляющая общения несет лишь 7 % общей информации, передаваемой между людьми при разговоре, в то время как интонация и выражение лица – 38 % и 55 % соответственно [1]. За последние несколько десятилетий получили развитие методы компьютерного распознавания эмоций, позволяющие повысить уровень взаимодействия между человеком и компьютером, а также в автоматическом режиме получить информацию об отношении человека к предметам или событиям.

Обзор методов распознавания эмоций

Большинство методов компьютерного распознавания эмоций основано на обучении с учителем. В качестве исходных данных в них используются либо отдельные изображения лица, либо последовательность кадров из видеопотока, в связи с чем методы можно условно разделить на статические и динамические. Динамические методы, как правило, используют информацию о движении лицевых мышц во времени. Так, например, один из самых первых методов автоматического распознавания эмоций использовал оптические потоки для определения направления и скорости движения отдельных частей лица, которые затем сопоставлялись с образцами для каждой из рассматриваемых эмоций [2]. Аналогичный подход встречается и в более поздних работах, однако вместо оптического потока чаще используется модель соединенных вибраций [3], позволяющая более точно определить деформации частей лица. Следует отметить, что существуют также наработки по автоматическому выделению паттернов. Например, авторы работы [4] используют скрытые марковские модели для автоматической сегментации аудио- и видеопоследовательностей с последующей разметкой полученных сегментов.

Одним из главных преимуществ динамических методов является то, что они позволяют уловить кратковременные изменения на человеческом лице – так называемые микроэмоции [5]. Однако во многих практических задачах (например, распознавание по одному изображению)

временная информация отсутствует. Поэтому также часто используются статические методы, основанные исключительно на визуальной оценке. Хотя существует множество способов представления лица (например, кривые Безье, используемые в работе [6]), наибольшее распространение при анализе эмоций получили модели активного образа [7–9]. Данный класс моделей позволяет эффективно вычислить положение ключевых точек, таких как центр зрачка глаза, уголки губ, контур носа и т. д., а затем уже на их основе построить распознавание. Наиболее подробный обзор методов, основанных на моделях активного образа, приведен в [9], где авторы сравнивают сразу 3 метода распознавания – евклидово расстояние, модели гауссовых смесей и метод опорных векторов.

Метод, описанный в данной работе, также является статическим и основан на анализе положения ключевых точек, полученных посредством модели активного образа, и наиболее близок к методу гауссовых смесей из [9], однако имеет несколько важных отличий. Во-первых, в приведенном в [9] методе в качестве признаков используется смещение ключевых точек, в то время как в данной работе используются расстояния между ними. Во-вторых, вместо гауссовой смеси для всего набора признаков используется набор функций распределения для каждого признака по отдельности, т. е. каждому признаку ставится в соответствие своя функция распределения. В-третьих, для определения вероятности появления признака используется куммулятивная функция вероятности.

В следующем разделе будет приведено общее описание метода, после чего кратко рассмотрены модели активного образа, а затем подробно описана разработанная статистическая модель распределения значений признаков и построенный на ней алгоритм распознавания.

Метод распознавания эмоций

Разработанный метод относится к классу методов обучения с учителем и состоит из двух этапов.

1. Этап обучения, на котором по исходным данным строится (обучается) модель распределения ключевых признаков.

2. Этап применения, на котором обученная модель применяется к реальным данным, в результате чего входные образцы относят к одному из заданных классов.

В качестве исходных данных для метода выступают размеченные вручную изображения человеческого лица: вместе с каждым изображением на вход алгоритма на этапе обучения поступает массив ключевых точек, описывающий наиболее характерные черты человеческого лица, а именно, брови, глаза, нос, губы, контур лица.

Первые 4 группы элементов, согласно [10], являются основными визуальными признаками, отражающими эмоциональное состояние человека, в то время как последний пункт является скорее техническим требованием для метода. Следует отметить, что контур лица не обязательно должен совпадать с контуром головы: как правило, для ограничения формы снизу добавляются точки на подбородке, сверху используются точки на бровях, а слева и справа контур просто замыкается (например, крайняя точка правой брови соединяется прямой линией с крайней правой точкой на подбородке).

Кроме этого, исходными данными являются также метки эмоций для каждого изображения. В данной работе используется ставший стандартом де факто набор из 6 базовых меток эмоций, описанный в [5] (удивление, радость, гнев, отвращение, печаль, презрение), а также специальная метка, обозначающая нейтральное выражение лица. Эта метка не присваивается входным изображениям, а вместо этого вычисляется на основе последнего компонента входных данных – порога вероятности «не нейтрального» выражения лица (процесс вычисления описан ниже).

В качестве признаков в разработанном методе используются расстояния между некоторыми парами ключевых точек, такими как центр глаза и центральная точка на брови, между точками в середине верхней и нижней губ и т. д. Все эти исходные данные используются для построения составной модели эмоций, состоящей из двух частей – модели активного образа, необходимой для получения ключевых точек, и статистической модели распределения значения признаков, используемой непосредственно для распознавания.

После того, как построены обе составляющие модели, они применяются к входным, неразмеченным изображениям. Вначале, с помощью модели активного образа на изображении ищутся ключевые точки. Затем, на основании положения этих точек автоматически вычисляются признаки – ключевые расстояния. Наконец, к вектору этих признаков применяется построенная модель распределения, в результате чего вычисляется вероятность принадлежности изображения к каждой из эмоций. Изображению присваивается метка той эмоции, которая получила наибольшую вероятность, либо метка нейтрального выражения лица, если ни одна вероятность не превысила входной порог.

Более подробно обе составляющие модели будут описаны в следующих разделах.

Модели активного образа

Прежде, чем начать разговор о моделях активного образа (active appearance models, ААМ [7, 8]), необходимо ввести два понятия.

1. Форма – набор точек на изображении, составляющих контуры изучаемого объекта. Например, форма кисти руки человека может быть задана через точки на контурах ладони и пальцев, а форма лица – через точки на бровях, носу, губах и внешнем контуре (рис. 1).

2. Текстура – интенсивность пикселей на изображении внутри внешнего контура формы.

Модели активного образа – это общее название нескольких алгоритмов для сопоставления статистических моделей формы и текстуры некоторого объекта с новым изображением. Следует отметить, что само название ААМ может относиться как ко всему алгоритму, так и к обученным моделям формы и текстуры (точное определение последних см. ниже). В контексте анализа эмоций ААМ служат для максимально точного поиска ключевых точек из статистической модели формы на изображении человеческого лица. Модель текстуры при этом используется как вспомогательный элемент при оптимизации («подгонке»).

Обучение модели активного образа происходит следующим образом [7]. Вначале все формы из обучающей выборки приводятся к единому размеру и углу наклона с помощью прокрустового анализа, что позволяет исключить ошибки масштабирования. Затем каждая форма рассматривается в новом пространстве признаков, где все точки формы образуют одну точку в многомерном пространстве. Поскольку положение многих точек на лице изменяется синхронно, возникает некоторая избыточность информации. Эту избыточность разрешают через анализ главных компонент: с помощью сингулярного разложения вычисляется матрица трансформации S , а вектора форм транслируются в пространство меньшей размерности.

Любая допустимая в рамках обучающей выборки форма в этом случае может быть выражена как

$$s = s_0 + s_p, \quad (1)$$

где s_0 – усредненная форма, а p – вектор параметров.

Уравнение (1) и выражает модель формы (shape model). Забегая вперед, можно сказать, что цель ААМ как раз и заключается в поиске значений вектора p .

Модель текстуры (appearance model) формулируется схожим образом: вначале каждая текстура переводится в один вектор в многомерном пространстве, а затем ко всему набору применяется МГК. После этого выводится соответствующая модель текстуры:

$$A(x) = A_0(x) + \sum_i \lambda_i \cdot A_i(x), \quad (2)$$

где A_0 – усредненная текстура, A_i – i -ый компонент (вектор-столбец) матрицы преобразования, а λ_i – соответствующий элемент вектора параметров модели.

Как уже было сказано выше, конечной целью моделей активного образа является поиск максимально точного соответствия между моделью и реальным изображением. Эта задача является задачей оптимизации и решается методом наименьших квадратов.

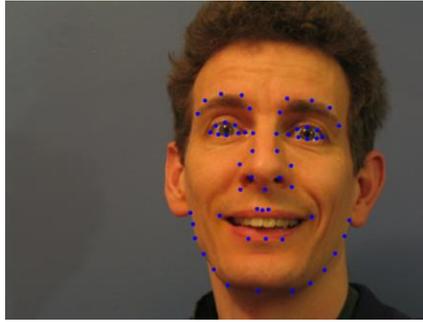


Рис. 1. Форма лица, заданная набором ключевых точек на изображении

Статистическая модель эмоций

По своей сути разработанная статистическая модель эмоций является вариацией наивного байесовского классификатора, где в качестве признаков используются расстояния между определенными ключевыми точками, например, между центральными точками верхней и нижней губ. Особенностью разработанного алгоритма также является то, что распределение вероятностей появления значений каждого признака по отдельности аппроксимируется графиком нормального распределения. Как и модели активного образа, модель эмоций имеет этап обучения и этап применения. Рассмотрим каждый из них по отдельности.

В качестве исходных данных при обучении модели используется набор форм, полученных при применении модели активного образа, а также соответствующие каждой форме метки класса. Если это не было сделано ранее, формы выравниваются к некоторой базовой (например, к усредненной), а также масштабируются до условного размера реального человеческого лица (в данной работе за высоту лица, т. е. расстояние между линией бровей и линией подбородка, была взята величина в 14 см).

При обучении модели для каждой пары признака и эмоции строится гистограмма, описывающая распределение значений этого признака. Например, на рис. 2 изображена гистограмма распределения расстояний между центром глаза и центральной точкой на брови при эмоции удивления. Высота каждого столбца на данной гистограмме отражает количество изображений людей, демонстрирующих удивление, для которых указанное расстояние попало в соответствующий промежуток значений. Данную гистограмму можно аппроксимировать графиком нормального распределения с центром в точке 2,1 см и стандартным отклонением 0,4 см. Следует отметить, что для обычного положения брови (без проявления каких либо эмоций) график выглядит примерно так же, но среднее значение сдвинуто в точку 1,2 см, а стандартное отклонение равно всего 0,25 см.

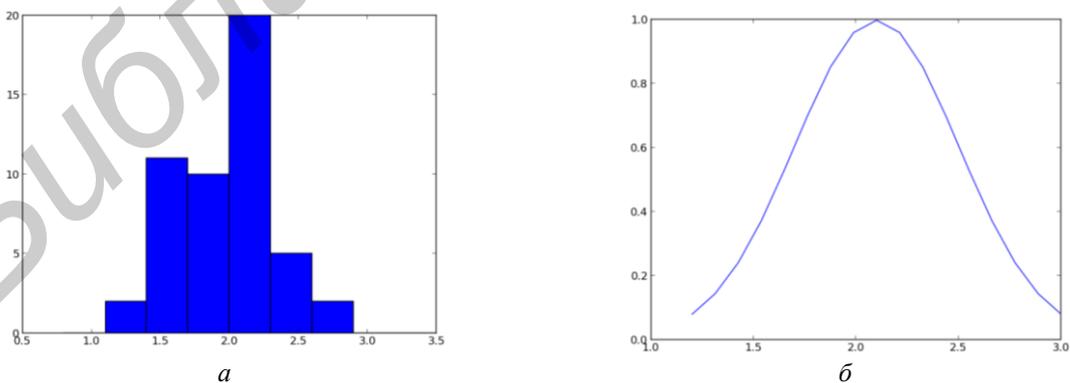


Рис. 2. Расстояния между центром глаза и центральной точкой на брови (в см):

a – гистограмма расстояний; b – график распределения, аппроксимирующего данную гистограмму

Современные математические и статистические пакеты, такие как R, позволяют не только аппроксимировать гистограмму графиком распределения вероятностей, но и вычислить параметры этого распределения. Зная параметры распределения, можно подсчитать вероятность появления данного признака при каждой из шести изучаемых эмоций. С

математической точки зрения, данная вероятность будет равна значению функции распределения (cumulative distribution function) в соответствующей точке графика. Так, например, на рис. 3 изображена функция плотности вероятности для рассматриваемого признака. Черная вертикальная линия соответствует значению признака в 1,7 см. Вероятность появления такого признака при эмоции удивления будет равна площади под графиком от его левой границы до черной линии, а именно, 0,159.

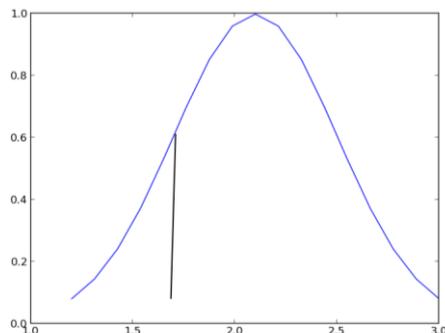


Рис. 3. Вычисление вероятности появления заданного значения признака

Строго говоря, между некоторыми признаками существует зависимость. Например, увеличение расстояния от центра глаза до брови всегда сопровождается увеличением расстояния между верхней и нижней веками. Однако точное моделирование всех внутренних зависимостей является сложной, если вообще реализуемой задачей, поэтому в данной работе используется «наивное» предположение о независимости признаков друг от друга. В этом случае можно рассчитать вероятность появления наблюдаемого набора признаков при каждой эмоции как произведение всех частных вероятностей:

$$P(x_1, x_2, \dots, x_n | E) = P(x_1 | E) \cdot P(x_2 | E) \cdot \dots \cdot P(x_n | E), \quad (3)$$

где x_1, x_2, \dots, x_n – признаки, E – наблюдаемая эмоция.

На выходе классификатора будет та эмоция, при которой вероятность наблюдения данного набора признаков максимальна, либо отсутствие эмоций, если полученное значение не превышает заданного порога:

$$e' = \arg \max_i \prod_i P(x_i | E), \quad (4)$$

$$e = \begin{cases} e' & \text{if } P(e') > t, \\ N & \text{otherwise,} \end{cases} \quad (5)$$

где e – итоговая эмоция, e' – предполагаемая эмоция, t – пороговое значение, а N – условный символ нейтрального выражения лица.

Следует обратить внимание, что в данном методе рассчитывается именно вероятность появления признаков при каждой из изучаемых эмоций, а не вероятность эмоции при наблюдаемых признаках. Для точного вычисления вероятности появления эмоции необходимо знать ее априорную вероятность, что на синтетических данных невозможно. Если же сделать предположение о равной априорной вероятности появления всех эмоций, то из теоремы Байеса очевидно, что значения $P(x_1, x_2, \dots, x_n | E)$ и $P(E | x_1, x_2, \dots, x_n)$ будут прямо пропорциональны.

Результаты и их обсуждение

Для проверки метода была использована библиотека FaceTracker [11]. Данная библиотека позволяет достаточно точно (хотя и не идеально) определять ключевые точки на лице человека, тем самым решая вопрос получения ключевых точек. В качестве признаков для составления статистической модели эмоций были использованы расстояния между определенными ключевыми точками, такими как расстояние между губами, крайними точками бровей, центром глаза и средней точкой на брови и др. Для обучения были использованы

фотоснимки из коллекции Пола Экмана [5] – хотя эта коллекция и не является наибольшей по размеру, она обладает тем преимуществом, что все изображения были отобраны Экманом как наиболее показательные для каждой из эмоций. Наилучшие результаты были получены при определении радости (87 %), наихудшие – для гнева (всего 62 %). Это можно объяснить слабой выраженностью этой эмоции в терминах ключевых точек, а также большой вариативностью ее выражения (главным признаком гнева являются сдвинутые брови, однако дисперсия расстояния между бровями очень велика; другие признаки могут выражаться несколькими способами, например, губы могут быть плотно сжаты или же, наоборот, раздвинуты, оголяя зубы). Точность классификатора на других примерах колебалась между значениями 69 и 81 %.

По мнению автора, низкая точность классификатора для многих примеров была вызвана тем, что многие эмоции имеют по несколько характерных «масок». Как уже было указано выше, гнев может быть выражен как плотно сжатыми, так и широко раскрытыми губами; то же самое можно сказать про отвращение, а для удивления Экман выделяет целых 4 возможных варианта. Данную проблему можно решить двумя способами: 1) создание отдельных статистических моделей для каждой «подэмоции»; 2) использование более сложных моделей распределения. Второй способ связан с вопросами вычисления функции распределения для нестандартного распределения, поэтому в будущем планируется в первую очередь апробировать первый подход. Другим направлением развития классификатора является улучшение модели активного образа для более точного определения ключевых точек, а также для включения дополнительных ориентиров, таких как морщины у глаз и губ (например, при улыбке).

Заключение

Разработан метод распознавания эмоций по изображению человеческого лица. В основе метода лежит статистическая модель распределения расстояний между ключевыми точками, описывающая возможные вариации элементов лица, а также позволяющая вычислить вероятность принадлежности текущего набора ключевых точек к тестируемой эмоции.

METHOD OF EMOTION RECOGNITION USING MODEL OF KEY DISTANCE DISTRIBUTION

A.V. ZHABINSKI

Abstract

A new method for emotion recognition from an image of person's face is presented. The method is based on our developed statistical model and probabilistic algorithm similar to naïve Bayes classifier. Training sets of key points are obtained using active appearance models.

Список литературы

1. Mehrabian A. // Psychology Today. 1968. Vol. 2 (11). P. 52.
2. Mase K. // IAPR Workshop on Machine Vision and Applications. Tokyo, 1990. P. 195–198.
3. Tao H., Huang T. S. // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, 23–25 June, 1998.
4. Cohen I., Huang T.S., Garg A. // Neural Information Processing Systems. 2000. Vol. 2.
5. Экман П. Психология лжи. СПб, 2009.
6. Khan M.I., Bhuiyan A. // International Journal of Computer Science and Network Security. 2009. № 9. P. 300–306.
7. Cootes T.F., Edwards G.J., Taylor C.J. // In Proc. European Conf. on Computer Vision. 1998. Vol. 2. P. 484–498.
8. Matthews I., Baker S. // International journal of computer vision. 2004. Vol. 60 (2). P. 135–164.
9. Ratliff M. S. Active appearance models for affect recognition using facial expressions: thesis by MD in CS. Wilmington, 2010.
10. Экман П., Дарвин Ч. О выражении эмоций у человека и животных. СПб, 2013.
11. Saragih: Главная страница. [Электронный ресурс]. – Режим доступа: <http://jsaragih.org>. – Дата доступа: 16.04.2013.