

СРАВНЕНИЕ СТАТИЧЕСКОГО И ДИНАМИЧЕСКОГО ПОДХОДОВ К КЛАССИФИКАЦИИ ЭМОЦИЙ ПО ВИДЕОИЗОБРАЖЕНИЮ ЛИЦА

Жабинский А. В.

Кафедра электронных вычислительных машин
Беларуский Государственный Университет Информатики и Радиоэлектроники
Минск, Республика Беларусь
E-mail: andrei.zhabinski@gmail.com

В данной работе проводится сравнение статического и динамического подходов к распознаванию эмоций по видеоизображению человеческого лица. В качестве практического базиса для сравнения предложено два метода, использующих, соответственно, один отдельный кадр и последовательность из двух кадров. Для получения признаков для распознавания в обоих методах использованы модели активного образа.

I. ВВЕДЕНИЕ

В последнее время взаимодействие человека с компьютером становится всё более и более полноценным: улучшаются алгоритмы распознавания речи и голосовые интерфейсы, появляются методы автоматического составления структурированных баз знаний, улучшаются алгоритмы поиска по запросам на естественном языке и т.д. Однако всё ещё существуют плохо изученные области такого взаимодействия и слабо формализованные задачи в них. Одной из таких задач является распознавание эмоций по человеческому лицу.

В данной работе рассмотрены статический и динамический подходы к распознаванию эмоций по набору изображений человеческого лица. При статическом подходе используются отдельные изображения, в то время как при динамическом – наборы (2 или более) последовательных кадров. В качестве алгоритма для получения признаков для распознавания используются модели активного образа, описанные в следующем разделе.

II. МОДЕЛИ АКТИВНОГО ОБРАЗА

Моделями активного образа (Active Appearance Models, ААМ)[1] называют класс алгоритмов компьютерного зрения для сопоставления статистических моделей формы и текстуры некоторого объекта с новым изображением. Под формой здесь подразумевается набор ключевых точек, описывающий основные контуры объекта, а под текстурой – интенсивность пикселей на изображении внутри самого внешнего контура. Для человеческого лица в качестве ключевых точек, как правило, используются точки на бровях, глазах, носу и губах, а также на внешнем контуре лица.

ААМ является алгоритмом обучения с учителем и, следовательно, разделён на 2 этапа: обучение и применение. На этапе обучения по набору размеченных изображений строятся статистические модели формы и текстуры (следует отметить, что в литературе по ААМ название

«модель активного образа» часто применяется не только ко всему алгоритму, но и конкретно к совокупности этих двух «подмоделей»). Обе модели являются параметрическими. Параметры модели формы контролируют отклонение положения ключевых точек от средней формы, а параметры модели текстуры, соответственно, – отклонение интенсивности пикселей от средней текстуры.

На этапе применения алгоритм старается найти такие значения параметров, при которых совокупная модель максимально точно соответствует новому изображению. Это означает, в частности, что алгоритм старается определить форму объекта на изображении, к которому применяется. Например, для модели человеческого лица выходными данными алгоритма являются положения ключевых точек на входном кадре.

В качестве метода оптимизации наибольшую популярность получил модифицированный метод Лукаса-Канаде с использованием обратного композиционного подхода [2].

Следует также отметить такую особенность ААМ, что вся работа с формами проводится после их предварительного выравнивания с помощью прокрустового анализа, поэтому в рамках метода можно считать их нормализованными и независимыми от масштаба и угла поворота.

III. СТАТИЧЕСКИЙ МЕТОД С ИСПОЛЬЗОВАНИЕМ КОНЕЧНОЙ ФОРМЫ

Принято считать, что эмоции на человеческом лице проявляются через его основные элементы – брови, глаза, нос и рот [3]. Именно эти элементы и описываются ключевыми точками формы, полученными после применения ААМ к изображениям лица. На основании этих координат можно построить следующий **статический** метод распознавания эмоций.

Вначале собирается набор данных, а именно изображения лиц и соответствующие им метки эмоций. Затем, к этим изображениям применяется ААМ для получения координат ключевых точек. После этого координаты организуются в

виде вектора вида $(x_1, y_1, x_2, y_2, \dots, x_n, y_n)^T$, где n – количество точек в модели формы. На основании набора таких векторов, а также вектора меток эмоций, обучается классификатор. В данной работе в качестве алгоритма классификации использовался метод опорных векторов (Support Vector Machine, SVM), хотя это и не является принципиальным.

После того, как классификатор построен, он применяется к новым изображениям по аналогии с этапом обучения: с помощью ААМ определяются координаты ключевых точек уже на тестовом изображении, которые затем переводятся в векторный вид и к ним применяется построенная модель SVM. Выходными данными здесь является метка распознанной эмоции.

Данный метод является статическим, поскольку не использует темпоральную информацию, а только отдельные изображения лиц.

IV. ДИНАМИЧЕСКИЙ МЕТОД С ИСПОЛЬЗОВАНИЕМ ПАРАМЕТРОВ ОПТИМИЗАЦИИ ААМ

Ещё одним источником информации для классификации являются вектора движения ключевых точек при проявлении эмоций. Например, при удивлении брови поднимаются вверх, а глаза расширяются. Поэтому возможно построить **динамический** метод распознавания, аналогичный описанному выше статическому, но с использованием в качестве признаков *вектора изменения параметров оптимизации модели формы*.

В качестве входных данных для одного объекта в данном методе используется два кадра, взятых с небольшим интервалом. Для простоты описания будем считать, что на первом кадре изображено лицо с нейтральным выражением, а на втором – с проявленной эмоцией. Тогда при оптимизации вектора параметров модели формы (назовём его p) ААМ выведет вектор изменения Δp , описывающий смещение ключевых точек относительно предыдущего положения. Набор именно таких векторов Δp затем используется для обучения классификатора и применения его к новым двухкадровым объектам. Выходными данными, как и в предыдущем случае, является метка распознанной эмоции.

V. СРАВНЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ

Для проверки производительности методов использовалась библиотека ICAAM [4], реализующая модели активного образа, а также изображения из неё же. Оба метода были опробованы на трёх эмоциях – радости, печали и удивлении. При этом эмоции сравнивались попарно, поскольку SVM по своей природе является бинарным классификатором. (Следует отметить, что существуют подходы, позволяющие использовать метод опорных векторов для классификации на большее количество классов, однако в целях исследования и для чистоты эксперимента имеет смысл применять SVM наиболее стандартным для него способом). Ниже приведена таблица сравнения производительности метода для каждой пары.

Таблица 1 – Сравнение методов

Эмоции	Стат. метод	Дин. метод
Радость/Печаль	0.778	0.667
Радость/Удивление	0.900	0.600
Печаль/Удивление	0.714	0.571

Как видно из таблицы, динамический метод показал гораздо более высокие результаты. И хотя данный тест нельзя считать всеобъемлющим, его результаты позволяют говорить о необходимости дальнейшего изучения динамического подхода к распознаванию эмоций.

VI. СПИСОК ЛИТЕРАТУРЫ

1. Cootes, T. F. Active appearance models / T. F. Cootes, G. J. Edwards, C. J. Taylor // In Proc. European Conf. on Computer Vision. – 1998. – Vol. 2, P 484–498.
2. Matthews, I., Baker, S. Active Appearance Models Revisited / I. Matthews, S. Baker // The Robotics Institute of Carnegie Mellon University. – 2004.
3. Экман, П. Психология лжи / П. Экман // Питер. – 2009.
4. ICAAM - Inverse Compositional Active Appearance Models [Electronic resource] / L. Vezzaro. – Mode of access: <http://www.mathworks.com/matlabcentral/fileexchange/32704-icaam-inverse-compositional-active-appearance-models>. – Date of access 10.09.2013.
5. 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.