

УДК 519.254

А.В. Жабинский

## РАСПОЗНАВАНИЕ ЭМОЦИЙ ПО ИЗОБРАЖЕНИЮ ЛИЦА С ИСПОЛЬЗОВАНИЕМ ТЕКСТУРНОГО АНАЛИЗА

*Рассматривается роль текстуры (интенсивности пикселей) в задаче распознавания эмоций по изображению человеческого лица. Проводится сравнение текстуры и ключевых точек, а также предлагается комбинированный метод распознавания, объединяющий оба набора признаков.*

### Введение

Понимание эмоционального состояния человека является ключевым элементом множества областей деятельности, таких как социологические и психологические исследования, маркетинг, игровая индустрия и др. В большинстве случаев эта информация передается от человека к человеку с помощью невербальных средств. По данным исследований [1] 55 % информации об эмоциях порождается выражением лица. Поэтому для полноценного взаимодействия между человеком и компьютером крайне важным является создание эффективных алгоритмов распознавания эмоций.

Несмотря на более чем двадцатилетнюю историю развития, область распознавания эмоций все еще является слабо формализованной. Так, разные методы используют разные типы информации (текстовую, визуальную, аудио и др.), разные типы признаков (статические и динамические), разные модели представления (оптические потоки, соединенные вибрации, ключевые точки и др.) и, конечно же, разные алгоритмы. Обзор основных направлений можно найти в [2]. В настоящей работе сосредоточимся на анализе статических изображений лиц.

Практически все подобные методы можно условно разделить на два шага: извлечение признаков и непосредственное распознавание, однако содержание этих шагов сильно варьируется.

Так, например, на первом шаге можно выделить две основные группы признаков: общевизуальные и специфичные для лиц. К первой группе можно отнести все популярные в компьютерном зрении признаки. В работе [3] для получения признаков используются фильтры Габора, а в [4] – локальные бинарные шаблоны. Эти методы легко реализовать, однако, как правило, полученные признаки дают относительно низкую точность распознавания. К признакам, специфичным для лиц, можно отнести, в частности, 46 единиц действий (англ. action units) из системы кодирования выражений FACS, описывающих положение лицевых мышц (см., например, [5]). Однако наибольшее распространение получили наборы ключевых точек, описывающих основные элементы лица, и метод их получения – так называемые модели активного образа [6–8]. Распознавание в этом случае сводится к получению координат этих точек и обучению классификатора на их основе. Подобные методы дают высокую степень распознавания, интуитивно понятны и согласуются с теорией психологии эмоций. Вместе с тем большинство таких методов игнорирует информацию об интенсивности пикселей в области лица при выражении эмоций (по причинам, которые будут описаны далее, совокупность этих пикселей будем называть текстурой). В данной работе проводится оценка значимости этой информации, а также предлагается комбинированный метод, повышающий точность распознавания по сравнению с классическим подходом.

### 1. Метод на основе положения ключевых точек

Для адекватной оценки влияния интенсивности пикселей в области лица на точность распознавания эмоций необходимо в первую очередь определить базовый алгоритм, с которым будет проводиться сравнение. В данной работе в качестве базового используется один из вариантов классического алгоритма на основе положения ключевых точек.

По сути, этот алгоритм разделен на два этапа: получение координат ключевых точек и классификация эмоций на их основе. Классификация объектов по вектору признаков (в данном

случае координат точек) является стандартной задачей обучения с учителем и не представляет интереса. Следует лишь отметить, что использовался один из наиболее популярных классификаторов, а именно метод опорных векторов (англ. support vector machines, SVM).

Получение ключевых точек, однако, является гораздо более сложной задачей (рис. 1). Публикации последних лет (см., например, [2]), а также специализация крупнейших лабораторий по данной тематике позволяют говорить о том, что наиболее популярным методом являются так называемые модели активного образа (англ. active appearance models, AAM). Вкратце опишем его суть.

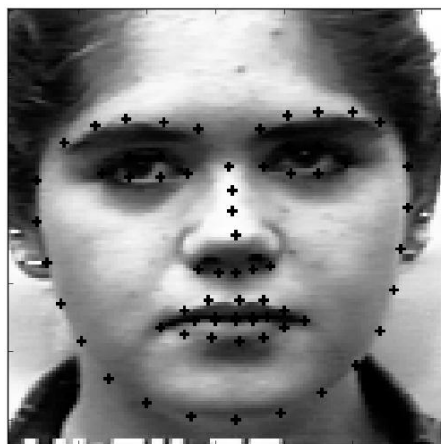


Рис. 1. Использование ключевых точек для описания контура основных элементов лица

Входными данными для метода является набор изображений лиц, в котором каждому изображению соответствует файл разметки, содержащий координаты ключевых точек, выбранных человеком. По этим данным ААМ строит две статистические модели:

*модель формы* – параметрическую линейную модель, описывающую возможные вариации положения ключевых точек. Формой при этом называется вектор координат ключевых точек  $s = (x_1, y_1, x_2, y_2, \dots, x_n, y_n)^T$ ;

*модель текстуры* – сходную модель, но описывающую уже возможные вариации интенсивности пикселей. Соответственно текстурой называется вектор всех пикселей внутри внешнего контура формы  $t = (t_1, t_2, \dots, t_n)^T$ .

Поскольку на разных изображениях количество пикселей внутри внешнего контура формы может быть разным, перед созданием модели текстуры все изображения лиц приводятся к единой усредненной форме посредством кусочно-аффинного преобразования: множество точек формы триангулируется, а затем каждый симплекс обычным аффинным преобразованием транслируется в новые координаты (рис. 2). На это преобразование следует обратить внимание, поскольку оно также будет использоваться в методе на основе информации о текстуре.

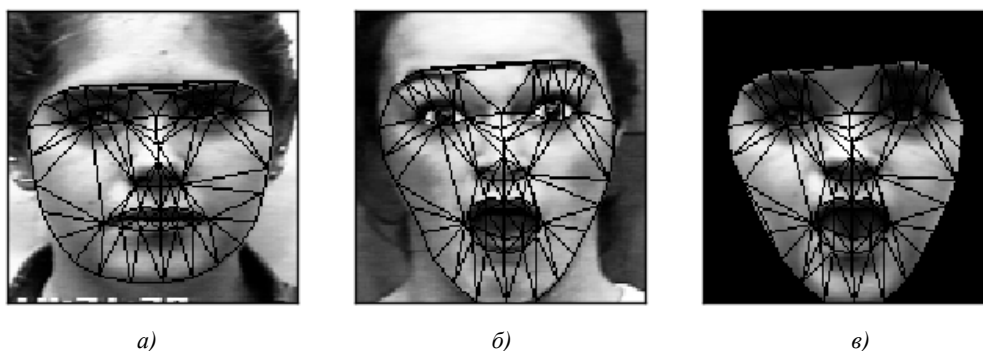


Рис. 2. Пример кусочно-аффинного преобразования: а) базовое изображение и соответствующая триангуляция по ключевым точкам (форме); б) аналогичное изображение с целевой формой; в) базовое изображение после приведения к целевой форме

При применении модели на новых изображениях вначале ищется примерное положение лица на фотографии, а затем полученные ранее модели формы и текстуры используются для итеративной подгонки под точные координаты элементов лица. Более подробно о моделях активного образа можно прочитать в [6, 7].

Здесь следует сделать замечание, которое является ключевым для данной работы: в этом подходе информация о текстуре используется для получения координат ключевых точек, но на этап классификации она не передается.

## 2. Метод на основе информации о текстуре

Возможность определения эмоций по положению и форме ключевых элементов лица хорошо изучена как психологами, так и специалистами в области информатики. В то же время аналогичные суждения в отношении текстуры практически не делались. Поэтому в первую очередь следует выдвинуть гипотезу о принципиальной возможности таких выводов.

**Гипотеза.** *Существует возможность определения (вручную или компьютерными средствами) выражения лица человека даже в отсутствие информации о положении и форме его ключевых элементов.*

Другими словами, предполагается, что эмоции на лице кодируются не только положением ключевых точек, но и другими признаками. Проверить это предположение, впрочем, не составляет труда. Для этого достаточно преобразовать лица, выражающие эмоции, к усредненной форме, а затем попытаться определить изначальное выражение. На рис. 3 изображен ряд лиц, к которым было применено такое преобразование.

Если присмотреться, то можно заметить, что все ключевые точки бровей, глаз, носа, губ, а также подбородка находятся в идентичном положении. Данные изображения получены путем триангуляции по ключевым точкам и применения уже знакомого нам кусочно-аффинного преобразования к снимкам лиц, выражающих эмоции. Несмотря на выравнивание положения ключевых элементов лица, эмоции на нем все еще легко угадываются по морщинам, теням и прочим элементам текстуры, что полностью подтверждает гипотезу.

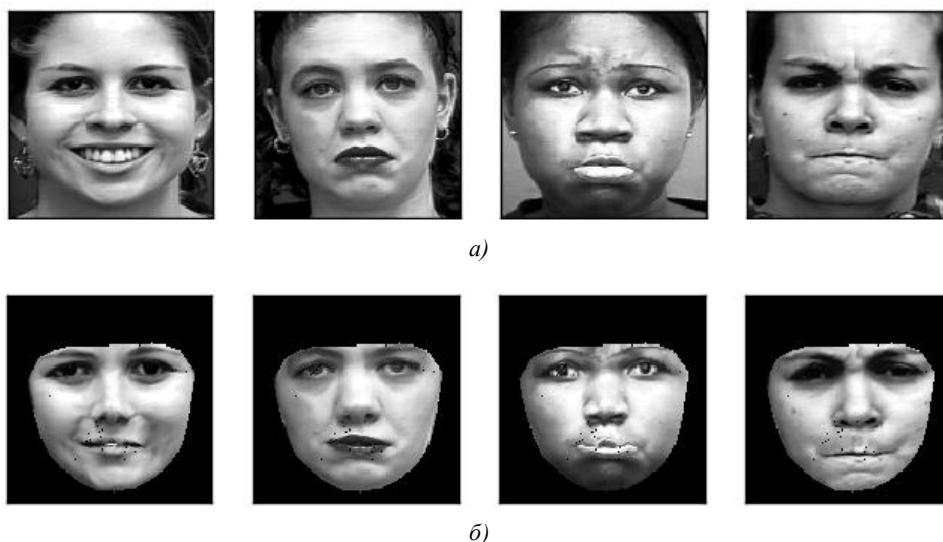


Рис. 3. Примеры приведения к средней форме: а) исходные изображения лиц; б) соответствующие изображения, приведенные к средней форме

Как и в методе на основе положения ключевых точек, преобразование к усредненной форме позволяет сформировать для каждого изображения соответствующий вектор текстуры:  $t = (t_1, t_2, \dots, t_n)^T$ , где  $n$  – количество пикселей внутри внешнего контура формы. Этот вектор можно напрямую использовать в качестве вектора признаков при классификации. В качестве классификатора можно применять уже известный нам SVM.

В итоге получаем следующий метод распознавания эмоций на основе информации о текстуре.

На этапе обучения:

1. Изображения подвергаются предварительной обработке, т. е. с помощью моделей активного образа на них определяются положения ключевых точек, а затем кусочно-аффинным преобразованием лица на изображениях приводятся к усредненной форме.

2. Пиксели полученных изображений вместе с соответствующими метками эмоций используются для обучения методом SVM.

На этапе применения:

1. По аналогии со стадией обучения новые изображения приводятся к средней форме.

2. Преобразованные изображения классифицируются с помощью обученной ранее модели.

Забегая вперед, скажем, что точность распознавания методом на основе текстуры лишь немного уступает точности классического подхода.

Подводя промежуточный итог, подчеркнем еще раз общие и отличительные черты двух методов. Оба включают этапы обучения (создания модели) и применения. Оба требуют предварительной обработки изображений для получения признаков. Первый метод использует в качестве признаков координаты ключевых точек (форму), полученных посредством моделей активного образа, второй – текстуру изображения внутри внешнего контура лица. В качестве конечного классификатора в обоих описанных методах используется SVM.

Также следует отметить, что в данной работе для приведения лиц к единой форме применяется модель активного образа, однако это не является обязательным требованием. Так, например, в работе [9] для аналогичных целей используется поиск нескольких ключевых точек и дальнейшее 3D-моделирование лица.

### 3. Комбинированный метод

Хотя описанные выше методы дают достаточно высокие результаты сами по себе, ничто не мешает объединить их, совместив оба набора признаков: координаты ключевых точек и интенсивности пикселей в области лица. Поскольку теоретическое обоснование для всех частей алгоритма уже было дано в предыдущих разделах, здесь просто опишем последовательность шагов комбинированного метода распознавания эмоций.

На этапе обучения:

1. К входным изображениям применяются модели активного образа для получения координат ключевых точек. Векторы ключевых точек для каждого изображения организуются в матрицу  $S$ .

2. К изображениям применяется кусочно-аффинное преобразование для приведения их к средней форме: для полученных на предыдущем шаге точек строится триангуляция Делоне, а затем пиксели внутри каждого отдельного треугольника отображаются на соответствующие пиксели внутри средней формы.

3. Пиксели каждого преобразованного изображения организуются в единый вектор признаков, а векторы для всех изображений организуются в матрицу  $T$ .

4. Матрицы  $S$  и  $T$  объединяются так, что каждый вектор  $s$  из матрицы  $S$  расширяется соответствующим вектором  $t$  из матрицы  $T$ . Другими словами, признаки из двух наборов объединяются в единый вектор. Результирующую матрицу назовем  $X$ .

5. Метки эмоций объединяются в единый вектор  $u$ .

6. Матрица данных  $X$  и вектор меток  $u$  используются для обучения модели SVM.

На этапе применения:

1. Новые изображения обрабатываются аналогично шагам 1–4 этапа обучения. В результате каждому изображению ставится в соответствие вектор  $x'$ , включающий в себя как данные о положении ключевых точек, так и пиксели трансформированного изображения.

2. Обученная на предыдущем этапе модель применяется к  $x'$  для предсказания изображенной эмоции.

Как и предполагалось, такой комбинированный подход дал точность, превосходящую результаты обоих описанных ранее методов.

#### 4. Результаты экспериментов

В качестве набора данных, на котором проводились эксперименты, использовалась расширенная база изображений Кона – Канаде (СК+) [10]. Эта база была составлена специально для изучения изображений лиц и в данный момент насчитывает 327 последовательностей кадров (от нейтрального выражения лица до ярко выраженной эмоции), при этом общее количество изображений превышает 10 000. Каждому изображению соответствует файл разметки с координатами ключевых точек  $x$ ,  $y$  (всего 68 пар), а каждой последовательности – файл с меткой эмоции. СК+ использует набор эмоций, предложенный Полом Экманом [11] и считающийся стандартом в задачах по их распознаванию. В этот набор входит шесть основных эмоций: гнев, страх, радость, грусть, удивление и отвращение. Дополнительно база СК+ содержит специальную метку отсутствия эмоций, однако в данной работе соответствующие изображения не использовались.

В качестве реализации ААМ была использована библиотека ААМToolbox [12], а для обучения конкретной модели активного образа – описанная выше база изображений СК+.

Полный цикл обучения моделей активного образа на базе данных СК+ на процессоре Intel i7 занимает около 3,5 ч. Время обучения классификатора SVM на 327 изображениях – от 5 до 7 с. Однако оба эти показателя относятся к стадии обучения. В то же время на стадии применения ААМ определяет положение ключевых точек в среднем за 180–190 мс, а SVM проводит классификацию примерно за 25 мс, что делает возможным использование разработанных алгоритмов в системах реального времени.

Точность распознавания определялась стандартным методом перекрестной проверки. При этом для обучения классификатора всегда использовалось последнее изображение (лицо, максимально сильно выражающее эмоцию) в каждой из последовательностей кадров (всего 327 изображений). По результатам десяти экспериментов были получены следующие результаты:

- метод на основе ключевых точек – 89,4 % правильных ответов;
- метод на основе текстуры – 86,3 % правильных ответов;
- комбинированный метод – 93,5 % правильных ответов.

Таким образом, использование информации о текстуре позволило достичь результатов, сравнимых с точностью распознавания по ключевым точкам (86,3 % против 89,4 % при классическом подходе). Наилучший же результат показал комбинированный метод, давший 93,5 % корректных ответов.

#### Заключение

На основе полученных результатов можно говорить о том, что текстура несет значительную информацию о выражении человеческого лица и ее использование целесообразно для повышения точности распознавания. Важным моментом является приведение изображений лиц к единой форме, без которого использование вектора текстуры было бы невозможным.

Несмотря на то что в данной работе использовалась непосредственно интенсивность пикселей, для классификации могут применяться и более сложные визуальные признаки (например, основанные на фильтрах).

#### Список литературы

1. Mehrabian, A. Communication without words / A. Mehrabian // Psychology Today. – 1968. – Vol. 2, no. 4. – P. 53–56.
2. Tian, Y. Facial expression recognition / Y. Tian, T. Kanade, J. F. Cohn // Handbook of face recognition. – London : Springer, 2011. – P. 487–519.
3. Littlewort, G. Dynamics of facial expression extracted automatically from video / G. Littlewort, M.S. Bartlett, I. Fasel // Image and Computer Vision. – 2004. – Vol. 24, no. 6. – P. 615–625.

4. Shana, C. Facial expression recognition based on Local Binary Patterns: A comprehensive study / C. Shana, S. Gong, P.W McOwan // Image and Vision Computing. – 2009. – Vol. 27, no. 6. – P. 803–816.
5. Senechal, T. Impact of action unit detection in automatic emotion recognition / T. Senechal, K. Bailly, L. Prevost // Pattern Analysis and Applications. – 2014. – Vol. 17, no. 1. – P. 51–67.
6. Cootes, T.F. Active appearance models / T.F. Cootes, G.J. Edwards, C.J. Taylor // Lecture Notes in Computer Science. – 1998. – Vol. 1407. – P. 484–498.
7. Matthews, I. Active appearance models revisited / I. Matthews, S. Baker. // Intern. J. of Computer Vision. – 2004. – Vol. 60, no. 2. – P. 135–164.
8. Ratliff, M.S. Active appearance models for affect recognition using facial expressions: thesis by MD in CS / M.S. Ratliff. – Wilmington, 2010. – Vol. 4. – 89 p.
9. DeepFace: closing the gap to human-level performance in face verification / Y. Taigman [et al.] // Conference in Computer Vision and Pattern Recognition. – Columbus, OH, USA, 2014. – P. 1701–1708.
10. Kanade, T. Comprehensive database for facial expression analysis / T. Kanade, J.F. Cohn, Y. Tian // Proc. of the Fourth IEEE Intern. Conf. on Automatic Face and Gesture Recognition. – Grenoble, France, 2000. – P. 46.
11. Lucey, P. The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression / P. Lucey, J.F. Cohn., T. Kanade // Proc. of the Third Intern. Workshop on CVPR for Human Communicative Behavior Analysis. – San Francisco, CA, USA, 2010. – P. 94–101.
12. Библиотека ААМToolbox [Электронный ресурс] – Режим доступа : <https://github.com/phoenix367/ААМToolbox>. – Дата доступа : 01.06.2014.

Поступила 03.06.2014

*Белорусский государственный университет  
информатики и радиоэлектроники,  
Минск, ул. П. Бровки, 6  
e-mail: andrei.zhabinski@gmail.com*

**A.V. Zhabinski**

## **EMOTION RECOGNITION USING TEXTURE ANALYSIS**

In the paper the role of texture (pixel values) in a task of emotion recognition in the image of a face is explored. Texture-based method is compared to classic methods based on coordinates of key point. In addition, a new combined method is presented that unites both sets of features.