

УДК 004.932.2

Сокращение размерности пространства *LBP*-признаков в задачах определения атрибутов личности по изображению лица

А. В. Рыбинцев*

Предложен подход к снижению размерности сформированного методом локальных бинарных шаблонов (locale binary patterns – LBP) пространства признаков, описывающих изображение за счет использования априорной информации о человеческом лице и алгоритма Adaboost для отбора значимых признаков. Приведены результаты вычислительных экспериментов, показывающие, что указанный подход снижает время классификации изображений почти в 8 раз.

Ключевые слова: компьютерное зрение, машинное обучение, классификация изображений, атрибуты личности, локальные бинарные шаблоны, отбор признаков.

Введение

Задача автоматического определения атрибутов личности по изображению лица на базе методов компьютерного зрения приобретает сегодня все большее значение, поскольку человек является наиболее часто встречающимся объектом на фотографиях и видео. Основными атрибутами личности, определение которых имеет первостепенное значение, являются пол человека, его раса и возраст [1, 2]. Среди практических задач, в которых именно эти атрибуты являются наиболее значимыми, следует выделить следующие:

- автоматическая нотация результатов поиска в Интернете;
- анализ аудитории при проведении маркетинговых мероприятий;
- формирование адаптивной наружной рекламы;

создание адаптивного человеко-машинного интерфейса;

контроль доступа личности к информации или определенным действиям.

Рассмотрим данные задачи более подробно.

В настоящее время интернет стал неотъемлемой частью нашей жизни, обойтись без которой не представляется возможным, а наличие целого ряда поисковых систем (Google, Yandex, Rambler и пр.) существенно упрощает процесс извлечения требуемой информации. Поскольку изображения людей часто являются предметом поиска или могут помочь в поиске необходимой информации, качество выполнения этой операции (количество найденных в сети объектов, действительно удовлетворяющих запросу) чрезвычайно важно. Если, например, необходимо найти изображение Иванова Сергея и известен год его рождения, то огромное количество изображений людей старшего возраста может быть отвергнуто поисковой системой, а если нужно найти изображение Сергиенко А. И. и известно, что это женщина, то все изображения мужчин так же не войдут в результат поиска.

* AVRybin@yandex.ru

При проведении маркетинговых или иных публичных мероприятий для лектора чрезвычайно полезно перед началом выступления оценить состав аудитории, так как ее возрастная, половая и расовая принадлежность могут существенно изменить набор используемых примеров, аналогий, цитат и шуток. Сделать это человеку достаточно трудно, тогда как автоматическая система, основанная на методах компьютерного зрения и установленная на входе в аудиторию, в состоянии учесть практически каждого человека.

Наружная динамическая реклама (огромные экраны, воспроизводящие рекламные ролики) постепенно вытесняет стационарные плакаты, так как ее эффективность намного выше. Объединение динамической рекламы с системой компьютерного зрения позволяет последней анализировать поток людей, проходящий рядом с таким экраном, с целью определения их атрибутов для выдачи наиболее значимой для данной категории лиц информации.

Поскольку компьютерные системы индивидуального обучения приобретают все большее распространение, чрезвычайно важно, чтобы процесс подачи учебного материала был максимально адаптирован к особенностям конкретной личности. При этом возраст, пол и даже раса человека могут иметь существенное значение для формирования темпа и порядка подачи учебного материала.

Приведенные выше примеры не требуют точной классификации людей по возрасту. Для принятия правильного решения достаточно определить его приближенно с точностью до 5 — 10 лет. Наряду с такими задачами существуют и другие, в которых требуется определять возраст с высокой точностью. Сегодня общепринята маркировка видеоматериалов по возрасту: 0+, 6+, 12+, 16+, 18+. Так, при автоматической продаже билетов на кинофильм категории 18+ целесообразно отказать 16-тилетнему подростку уже на этапе продажи билета, а не в момент его входа в кинозал. Подобный контроль необходим и при доступе к видеoinформации, размещенной в Интернете (сегодня практически все устройства доступа имеют встроенную видекамеру, так что технически эта задача вполне выполнима).

Технологический процесс определения атрибутов личности по изображению лица состоит из трех этапов [3 — 5]:

- нормализация изображения;
- формирования вектора признаков, описывающего изображение;
- классификации изображения.

В процессе нормализации изображение приводится к некоторому унифицированному виду, допускающему применение единых методов его обработки на последующих этапах. Данный процесс предусматривает:

- распознавание лица и его выделение из общего изображения;
- преобразование изображения из цветного в одноканальное (оттенки серого);

выделение на изображении ключевых точек (как правило — глаз) и выравнивание его относительно линии уровня глаз;

обрезание изображения, исходя из среднестатистических пропорций лица человека на основе расстояния между зрачками;

понижение уровня шума в исходном изображении за счет проведения медианной фильтрации [6], позволяющей убрать резкие выбросы, что весьма актуально для реальных фотографий;

получение фронтального изображения лица из реального за счет применения афинного преобразования или деформации на основе триангуляции Делоне [7].

Перечисленные процедуры, выполняемые в процессе нормализации изображения, являются достаточно сложными, а их совершенствование может быть предметом отдельного исследования. В настоящей работе предполагается, что этап нормализации уже выполнен, и исследуется следующий этап технологического процесса — формирование вектора признаков.

Признаковое описание изображения

Для определения атрибутов личности по нормализованному изображению лица с помощью математических методов необходимо иметь формальное описание исходного изображения. Таким описанием является вектор признаков (числовых характеристик) изображения. На этапе создания вектора признаков (в англоязычной литературе используется термин — features extraction), формируется пространство признаков и определяются их численные значения, т.е. применяется некий метод, меняющий реальное изображение на набор численных значений характеризующих его признаков.

При использовании методов компьютерного зрения для определения атрибутов личности по изображению лица в процессе признакового описания нормализованных изображений I применяются различные методы Ψ формирования вектора признаков X . Их можно разделить на две большие группы [4]:

методы Ψ_f , основанные на использовании текстуры (значений интенсивности пикселей) изображения;

методы Ψ_a , основанные на выделении антропометрических точек на изображении лица человека с последующим выделением информации о расстоянии между этими точками и их взаимном расположении.

В последние годы наибольшее распространение в рамках группы Ψ_f получил метод локальных бинарных шаблонов (Local Binary Patterns, LBP) [8 — 12], что объясняется целым рядом важных особенностей получаемого с его помощью вектора признаков. Локальный бинарный шаблон — это определенный вид признака, представляющий собой описание окрестности пиксела изображения в двоичном представлении. Базовый оператор LBP , применяемый к пикселу, использует восемь пикселей окрестности, принимая значение интенсивности центрального пиксела в качестве порога. Пиксе-

лы со значением интенсивности большим или равным значению интенсивности центрального пиксела принимают значения, равные «1», остальные — равные «0». Таким образом, итогом применения базового оператора *LBP* к пикселу изображения является восьмиразрядный бинарный код, описывающий окрестность этого пиксела (запись производится по часовой стрелке, начиная с верхнего левого пиксела). Далее этот код рассматривается как двоичная запись некоторого числа, и именно оно и ставится в соответствие пикселу.

После получения для каждого пиксела значения *LBP* строится гистограмма значений, т.е. формируется множество частот \bar{x}_i появления бинарных шаблонов в изображении i : $\bar{x}_i \in \mathbb{R}$; $|\mathbf{X}| = 256$; $0 < \bar{x}_i < 1$. Изображения лиц могут рассматриваться как набор локальных особенностей изображения, хорошо описывающихся с помощью локальных бинарных шаблонов:

$$\{x_i\} = (x_{i1}, x_{i2}, \dots, x_{i256}).$$

Однако гистограмма, построенная для всего изображения в целом, кодирует лишь наличие тех или иных локальных особенностей, но при этом не содержит никакой информации об их расположении на изображении. Для учета такого рода данных нормализованное изображение разбивается на регионы, для каждого из которых вычисляется своя гистограмма *LBP*, т.е. для региона j изображения i будет сформирована гистограмма $\{x_i^j\}$ значений *LBP*:

$$\{x_i^j\} = (x_{i1}^j, x_{i2}^j, \dots, x_{i256}^j).$$

Результирующим описанием изображения i является вектор значений \bar{x}_i , сформированный как конкатенация гистограмм, полученных по различным регионам исходного изображения:

$$\bar{x}_i = (x_{i1}^1, x_{i2}^1, \dots, x_{i256}^1, x_{i1}^2, x_{i2}^2, \dots, x_{i256}^2, \dots, x_{i1}^K, x_{i2}^K, \dots, x_{i256}^K),$$

где K — число регионов на которое разделено изображение.

Исходя из соотношений геометрических размеров характерных элементов лица человека (глаз, носа, рта, бровей и пр.) нормализованное изображение традиционно разбивается на $(6 \times 7 = 42)$ области [8].

Благодаря особенностям применения оператора *LBP*, локальные бинарные шаблоны инвариантны к небольшим изменениям в условиях освещенности и поворотам классифицируемого изображения, что и обусловило их широкое распространение для решения задач определения таких атрибутов личности, как «пол» [12 — 15] и «раса» [16 — 18]. Примеров использования метода *LBP* для определения точного значения атрибута «возраст» методами восстановления регрессии в доступной литературе не обнаружено. Основной причиной этого является очень большая размерность пространства признаков, формируемого данным методом. Так при разбиении изображения на $6 \times 7 = 42$

блока размерность пространства признаков $|\bar{x}_i| = 42 \times 256 = 10752$.

В работе [19] показано, что не все бинарные шаблоны обладают одинаковой информативностью. Наиболее важными являются равномерные шаблоны, к которым относятся бинарные комбинации, содержащие не более трех серий «0» и «1» (например, 00000000, 001110000 и 11100001). Равномерные *LBP* определяют только существенные локальные особенности изображения: концы линий, грани, углы и пятна. Их использование приводит к существенному уменьшению размерности пространства признаков, так как в этом случае используется только 58 шаблонов вместо 256 (поскольку шаблоны имеют разрядность $p = 8$, то общее количество таких шаблонов равно 2^p , т.е. 256, а количество равномерных шаблонов равно $p(p - 1) + 2 = 58$ [19]). В этом случае при разбиении изображения на 42 блока размерность пространства признаков равна $42 \times 58 = 2436$, что почти в 4,5 раза меньше исходного, но все равно остается весьма большой.

Итогом применения оператора *LBP* при использовании разбиения изображения на 42 блока и использовании равномерных шаблонов становится следующий вектор признаков для изображения i :

$$\bar{x}_i = (x_{i1}^1, x_{i2}^1, \dots, x_{i58}^1, x_{i1}^2, x_{i2}^2, \dots, x_{i58}^2, \dots, x_{i1}^{42}, x_{i2}^{42}, \dots, x_{i58}^{42}).$$

С одной стороны, столь высокая размерность вектора признаков является важным достоинством, так как подобное признаковое описание содержит много информации об исходном изображении. Но с другой, — использование вектора признаков такой размерности существенно усложняет задачу классификации. В [20] равномерные *LBP* использованы для определения значения возраста путем многоклассовой классификации (возрастная группа сжата до одного года, что позволяет определить точное значение возраста), но затраты времени на классификацию весьма велики.

Метод снижения размерности пространства признаков

Основным недостатком метода *LBP* является очень высокая размерность пространства формируемых им признаков (около 2500). Рассмотрим возможности снижения размерности данного пространства за счет использования априорной информации об изображении лица человека. В качестве исходных предпосылок для успешного решения этой задачи выделим следующие: симметричность лица;

различная информативность отдельных участков изображения лица.

Первая предпосылка весьма очевидна, однако использование эданного факта в литературе не обнаружено. При анализе алгоритма формирования пространства признаков методом *LBP* легко заметить, что итоговая гистограмма будет содержать практически

идентичные пары гистограмм, построенные по симметричным участкам лица. Поэтому, удаление половины изображения (половины гистограмм) никак не повлияет на информативность вектора признаков (трудно предположить, что пол, раса или возраст левой половины изображения лица будут отличаться от значения этих атрибутов для правой половины). Использование данного очевидного факта приводит к снижению размерности пространства признаков в 2 раза. Тогда при рассмотрении только равномерных бинарных шаблонов и разбиении исходного нормализованного изображения на 42 региона размерность пространства признаков снижается с 2436 до 1218, так как число регионов равно 21. В этом случае вектор признаков изображения \bar{x}_i ($i = 1, n$) принимает вид (с учетом использования равномерных бинарных шаблонов):

$$\bar{x}_i = (x_{i1}^1, x_{i2}^1, \dots, x_{i58}^1, x_{i1}^2, x_{i2}^2, \dots, x_{i58}^2, \dots, x_{i1}^{21}, x_{i2}^{21}, \dots, x_{i58}^{21}),$$

где второй нижний индекс (1, 2 ... 58) определяет номер равномерного шаблона, а верхний индекс — номер региона изображения.

Следует отметить следующее дополнительное преимущество использования априорной информации о симметричности лица. Как отмечалось ранее, на этапе нормализации изображение лица подвергается аффинному преобразованию с целью получения фронтального изображения из реального, которое практически всегда несколько повернуто. При этом наибольшие трудности с качеством изображения возникают при восстановлении дальней (частично скрытой) половины лица. Предлагаемая идея использования только одной половины изображения делает вообще не нужной процедуру восстановления второй половины, что сокращает затраты времени на процесс нормализации изображения без снижения качества результата.

Вторая предпосылка к снижению размерности пространства признаков (различная информативность участков изображения лица) так же представляется вполне очевидной (например, регион изображения, содержащий глаз или рот гораздо более информативен, чем регион, содержащий щеку). Да и внутри региона различные участки изображения содержат разное количество информации, скрытой в бинарных шаблонах (частично данный факт отражен в равномерных бинарных шаблонах). Поэтому из итогового вектора признаков желательно удалить те признаки, которые не оказывают влияния на итоговый результат классификации.

Пусть, по прежнему, имеется обучающая выборка по атрибуту «пол» $\mathbf{X}' \subset \mathbf{X}$, $|\mathbf{X}'| = n$, заданная множеством пар прецедентов $\{\bar{x}_i, y_i\}$, $i = 1, n$; $\bar{x}_i \in \mathbb{R}^m$, $y_i \in \{-1, +1\}$, где вектор признаков изображения \bar{x}_i получен по половине исходного изображения. Для простоты обозначения пронумеруем все признаки от 1 до 1218:

$$\bar{x}_i = (x_i^1, x_i^2, x_i^3, \dots, x_i^{1218}).$$

Для автоматического решения данной задачи: выделения наиболее информативных признаков из исходных, число которых равно 1218, предлагается использовать алгоритм Adaboost для бинарной классификации по атрибуту «пол». Основная идея заключается в том, чтобы применить процедуру бустинга к каждому из признаков, т.е. заставить алгоритм Adaboost принимать решение об отнесении изображения к тому или иному классу по каждому единичному признаку x_i^j отдельно, $i = (1, n)$; $j = (1, 1218)$. В этом случае задача классификации с помощью алгоритма Adaboost формулируется следующим образом:

$$f(x^j) = \text{sign}\left(\sum_{i=1}^n \alpha^j b(x_i^j)\right),$$

т.е. решающая функция f будет сформирована таким образом, чтобы классифицировать изображения только по j -му признаку ($j = 1, 1218$). В данном соотношении весовые коэффициенты α^j определяют значимость базовой функции $b(x)$, в качестве которой предлагается использовать простейшие бинарные классификаторы с разделением по пороговому значению (Decision Stump). Данное пороговое значение вычисляется, как среднее значение параметра по обучающей выборке.

Функционал качества $Q(f(x^j))$ определим как число ошибок, допускаемых комитетом из $k = 1218$ решающих базовых функций (по числу признаков) на признаках x^j всех объектов обучающей выборки \mathbf{X}' :

$$Q(f(x^j)) = \sum_{i=1}^n \alpha^j b(x_i^j) y_i < 0.$$

В соответствии с алгоритмом Adaboost вместо пороговой функции $z < 0$ будем использовать ее непрерывно дифференцируемую оценку сверху вида $E(z) = \exp(-z)$:

$$Q(f(x^j)) \leq \tilde{Q}(f(x^j)) = \sum_{i=1}^n \exp(-\alpha^j b(x_i^j) y_i).$$

Минимизация данного функционала приводит к следующему результату для весовых коэффициентов [21]:

$$\alpha^j = \frac{1}{2} \ln \frac{P(x^j)}{N(x^j)},$$

где $P(x^j)$ — доля правильных классификаций; $N(x^j)$ — доля ошибочных классификаций по признаку x^j на всех объектах обучающей выборки: $P(x^j) > N(x^j)$, $P(x^j) + N(x^j) = 1$. Для признаков x^j , у которых $P(x^j) < N(x^j)$, коэффициенты $\alpha^j = 0$. Полученные весовые коэффициенты нормируются так, чтобы их сумма была равна 1.

Результатом применения данного подхода будет совокупность весовых коэффициентов α^j , определяющих значимость базового алгоритма $b(x)$. Поскольку базо-

вый алгоритм применялся к единственному признаку x^j , данный весовой коэффициент будет характеризовать значимость конкретного признака x^j при принятии решения об отнесении объекта к определенному классу.

Далее предлагается упорядочить весовые коэффициенты по убыванию, и ввести порог значимости δ . Для решения задачи классификации предлагается использовать только первые N признаков, для которых выполняется условие:

$$\sum_{j=1}^N \alpha^j < \delta.$$

Полученные признаки будем называть усиленными. Таким образом, в каждом регионе изображения лица будут оставлены только те бинарные шаблоны, суммарная значимость которых выше заданного порога.

Результаты численных экспериментов

Результаты численных экспериментов, проведенные на существующей открытой базе изображений MORTH [22], содержащей 5493 изображения лица людей, показали, что предлагаемый подход, даже при значении $\delta = 0,99$ (оставляем признаки, имеющие хоть какую-нибудь значимость для классификации), позволяет отбросить более половины признаков и сократить размерность пространства признаков примерно до 480, т.е. почти в 5 раз по сравнению с традиционным использованием равномерных *LBP*, а для порога значимости $\delta = 0,95$ число усиленных признаков сокращается до 280, что почти на порядок меньше исходного.

Столь существенное сокращение размерности пространства признаков позволяет использовать метод *LBP* для решения задачи классификации изображений по атрибуту «возраст». Для сравнения были использованы результаты работы [20], где для возрастной классификации формировалось множество из K (по числу классифицируемых возрастов) бинарных классификаторов, каждый из которых определял, старше ли анализируемый человек конкретного возраста или нет. Для определения возраста человека нужно применить все K классификаторов, после чего итоговая оценка возраста считается равной числу положительных откликов всех классификаторов плюс один. В качестве критерия качества классификации традиционно использовалась мера *MAE* (Mean Absolute Error), определяющаяся следующим образом:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - f(\bar{x}_i)|.$$

Данная мера характеризует среднее абсолютное отклонение предсказанных решающей функцией f значений искомого атрибута для объекта \bar{x}_i от истинных значений этого атрибута для данного объекта; n — количество объектов в тестовой выборке.

В таблице приведено сравнение результатов, полученных в [20], с результатами, полученными аналогичным способом на идентичном компьютере, но только при использовании усиленных *LBP*, сформированных предлагаемым в настоящей работе методом. Отметим, что в [20] для сокращения времени классификации предлагалось использовать ряд упрощений в базовом алгоритме, что приводило к снижению точности классификации. Использование усиленных *LBP* позволило добиться аналогичного сокращения времени классификации практически без потери точности.

Заключение

В настоящей статье предложен подход к сокращению размерности пространства признаков изображения лица человека, сформированного методом *LBP*, за счет использования априорной информации о лице и алгоритма Adaboost. Полученное в результате пространство усиленных признаков имеет примерно на порядок меньшую размерность по сравнению со стандартным методом *LBP*, а использование усиленных признаков приводит к пропорциональному сокращению времени классификации (не только по полу, но и по возрасту) изображений лиц практически без потери точности. Следовательно, использование усиленных *LBP* в задачах определения атрибутов личности по изображению лица представляется целесообразным.

Работа выполнена при поддержке РФФИ (проект № 14-01-00671а).

Литература

1. Коршунов А. и др. Определение демографических атрибутов пользователей микроблогов // Труды Института системного программирования РАН. 2013. Т. 25. С. 179 — 194.
2. Куликова А.А. Подход к классификации пользователей в социальных сетях // Восточно-Европейский журнал передовых технологий. 2011. № 2 (51). С. 246—452.

Сравнение результатов возрастной классификации на основе *LBP* и усиленных *LBP*

	Равномерные <i>LBP</i> $N = 2436 (58 \times 42)$	Половина изображения $N = 1218 (58 \times 21)$	Усиленные <i>LBP</i> $N = 460 (\delta = 0,99)$	Усиленные <i>LBP</i> $N = 280 (\delta = 0,95)$
MAE	4,52 [20]	4,53	4,53	4,55
Время классификации, с	3,21 [20]	1,63	0,69	0,42

3. **Qiu X., Sun Zh., Tan T.** Global Texture Analysis of Iris Images for Ethnic Classification // Lecture Notes in Computer Sci. 2005. V. 3832. P. 411 — 418.
4. **Fu Y., Xu Y., Huang T.S.** Estimating human ages by manifold analysis of face pictures and regression on aging features // Proc. IEEE Conf. multimedia and Expo. 2007. P. 1383 — 1386
5. **Ramanathan N., Chellappa R.** Modeling age progression in young faces // Proc. IEEE Conf. computer vision and pattern recognition (CVPR'06). 2006. P. 387 — 394.
6. **Хуанг Т.С. и др.** Быстрые алгоритмы в цифровой обработке изображений. М.: Радио и связь, 1984.
7. **Yang Z., Li M., Ai H.** An experimental study on automatic face gender classification // Proc. Intern. Conf. Pattern Recognition (ICPR). 2006. P. 1099 — 1102.
8. **Shan C.** Learning local binary patterns for gender classification on real-world face images // Pattern Recognition Lett. 2012. V. 33 (4). P. 431 — 437.
9. **Hadid A., Pietikainen M.** Combining appearance and motion for face and gender recognition from videos // Pattern Recognition Lett. 2009. V. 42 (11). P. 2818 — 2827.
10. **Маенраа Т.** The Local binary pattern approach to texture analysis — extensions and applications. Oulu University Press, 2003.
11. **Yilionias J., Hadid A., Hong X., Pietikainen M.** Age estimation using local binary patterns kernel density estimate // Proc. Intern. Conf. image analysis and processing (ICIAP'13). 2013. P. 141 — 150.
12. **Lian H., Lu B.** Multi-view gender classification using local binary patterns and support vector machines // Proc. Intern. Symp. on neural networks. 2006. P. 202 — 209.
13. **Alexandre L.A.** Gender recognition: a multiscale decision fusion approach // Pattern recognition Lett. 2010. V. 31. P. 1422 — 1427.
14. **Chen C., Ross A.** Evaluation of gender classification methods on thermal and near-infrared face images // Proc. Intern. joint Conf. on biometrics (IJCB). 2011. pp. 367 — 374.
15. **Makinen E., Raisamo R.** An experimental comparison of gender classification methods // Pattern recognition Lett. 2008. V. 29. P. 1544 — 1556.
16. **Bellustin N. et al.** Instant Human Face Attributes Recognition System // Intern. J. Advanced Computer Sci. and Appl.: Special Issue on Artificial Intelligence. 2011. P. 112 — 120.
17. **Gutta S., Huang J., Jonathon P., Wechsler H.** Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces // IEEE Trans. on neural networks. 2000. V. 11. N 4. P. 948 — 960.
18. **Yang Z., Ai H.** Demographic classification with local binary patterns // Proc. IEEE Intern. Conf. biometrics. 2007. P. 464 — 473.
19. **Ojala T., Pietikainen M., Maenraa T.** Multiresolution gray-scale and rotation invariant texture classification with local binary patterns // IEEE Trans. on pattern analysis and machine intelligence. 2002. V. 24 (7). P. 971 — 987.
20. **Рыбинцев А.В., Лукина Т.М., Конушин В.С., Конушин А.С.** Возрастная классификация людей по изображению лица на основе метода ранжирования и локальных бинарных шаблонов // Системы и средства информатики. 2013. Т. 23. № 2. С. 62 — 73
21. **Вьюгин В.В.** Математические основы теории машинного обучения и прогнозирования. М.: МЦМНО, 2013.
22. **MORPH** (Craniofacial Longitudinal Morphological Face Database) [Электрон. ресурс]. <http://www.faceaginggroup.com/morph/> (дата обращения 01.12.2015).

Статья поступила в редакцию 06.11.2015