

Алгоритм приближенного поиска ближайшего соседа в задаче распознавания лиц с использованием опорных эталонов и статистического подхода

Савченко А. В., Chaudhury S.,
avsavchenko@hse.ru

Аннотация

Исследована задача распознавания лиц при наличии большого числа классов и малого количества эталонных фотографий для каждого класса. Показано, что современные методы решения задачи, основанные на сопоставлении векторов признаков высокой размерности, выделенных с помощью глубоких сверточных нейронных сетей, характеризуются низкой вычислительной эффективностью. Предложен алгоритм приближенного поиска ближайшего соседа, в котором следующий проверяемый эталон выбирается так, чтобы максимизировать условную плотность вероятности расстояний до эталонных изображений, проверенных на предыдущих этапах. При этом для снижения требования к дополнительной оперативной памяти вычисляются расстояния от всех изображений обучающего множества до небольшого количества опорных эталонов. В рамках экспериментального исследования для баз данных фотографий лиц LFW и PubFig83 показано, что предлагаемый алгоритм позволяет в 1,4-4 раза ускорить процедуру распознавания по сравнению с известными методами приближенного ближайшего соседа.

Ключевые слова: Распознавание лиц, методы приближенного ближайшего соседа, статистическое распознавание образов, глубокие нейронные сети.

1. ВВЕДЕНИЕ

Среди практически важных задач распознавания образов в последние годы повышенное внимание со стороны исследователей и практиков уделяется задаче классификации (или идентификации) людей по фотографиям их лиц [1]. С одной стороны, распознавание лиц является одним из наиболее сложных приложений методов классификации, с другой, сказывается бурный рост спроса на автоматические системы видеоконтроля и видеонаблюдения. Современные информационные технологии видеоидентификации людей по их фотографиям признаны одними из наиболее приемлемых для массового применения, так как они не требуют физического контакта объекта (человека) с устройством наблюдения и, в потенциале, характеризуется высокой надежностью ввиду известной уникальности лица каждой отдельной личности. Рынок программных продуктов распознавания лиц весьма разнообразен, при этом для массового потребления доступен целый ряд работающих технологий. Однако интенсивность исследований в этой области отнюдь не снижается, практические потребности в них только нарастают, что связано, в первую очередь, с вариативностью зрительных образов. Для достижения приемлемой точности классификации требуется реализовать сложные (с вычислительной точки зрения) алгоритмы, которые требуют наличия для каждого класса (человека) нескольких десятков фотографий, выполненных при разных условиях (освещения, поза, эмоции). Увеличение же числа фотографий, даже если оно возможно для конкретной прикладной задачи, приводит к резкому росту объема базы

данных, что, в свою очередь, еще более усложняет алгоритмы распознавания. В результате работающие технологии либо реализуются в виде сервиса (с очевидными недостатками – требованием к постоянному Интернет-соединению, возможными проблемами безопасности) или накладывают ограничение на максимальное количество классов для распознавания. Например, в текущей версии Intel RealSense SDK поддерживается распознавание лиц для не более, чем 20 пользователей. Таким образом, чрезвычайно актуальной является задача повышения вычислительной эффективности распознавания лиц при приемлемой точности для баз данных, включающих тысячи альтернативных классов.

2. ОПИСАНИЕ АЛГОРИТМА

2.1 Постановка задачи

Задача распознавания лиц состоит в том, чтобы поступающему на вход изображению лица X ставится в соответствие один из $C > 1$ заранее точно неопределенных классов (идентификатор человека). Предполагается, что для обучения системы доступна база данных, содержащая C эталонных изображений лиц с известным идентификатором распознаваемых людей (меткой класса) для каждого r -го эталонного изображений.

Предположим далее, что границы лиц были выделены с помощью известных алгоритмов (например, метод Виоль-Джонса), поэтому все изображения содержат только области лиц. Как известно, во многих промышленных системах проявляется проблема *малых выборок* (N): число эталонов для каждого класса недостаточно для обучения сложного классификатора. В таком случае обычно применяются методы ближайшего соседа – решение принимается в пользу класса r , где

$$v = \underset{r \in \{1, \dots, R\}}{\operatorname{argmin}} \rho(X, X_r), \quad (1)$$

Для вычисления меры близости $\rho(X, X_r)$ изображениям X и

ставится в соответствие вектор признаков размера M . В ранних исследованиях применялись традиционные признаки, например, гистограммы ориентированных градиентов и/или локальных бинарных шаблонов. В настоящий момент наиболее часто для настройки классификатора при распознавании лиц используется не доступное обучающее множество, а большая коллекция собранных в Интернете фотографий лиц, которая используется для обучения глубокой сверточной нейронной сети [5]. Далее значения на выходе предпоследнего слоя нейронной сети используются в качестве множества признаков любой фотографии лица, т.е. обученная сеть используется как способ извлечения признаков. Такой подход позволяет использовать накопленные большие объемы визуальной информации даже в случае малых обучающих выборок (N).

Алгоритмическая сложность реализации подхода (1), основанного на полном переборе множества эталонов,

составляет $O(R \cdot M)$. В результате, в связи с большой размерностью пространства признаков M , необходимых для описания лица, методы ближайшего соседа типа (1) оказываются неприемлемыми для реализации многих систем биометрической идентификации реального времени уже при наличии сотни классов. В таком случае для оценки эффективности методов распознавания следует учитывать не только среднюю вероятность ошибки классификатора $\bar{\alpha}$, но накладывать ограничение на допустимое среднее время принятия решений \bar{t} . Рассмотрим в следующем разделе способы повышения вычислительной эффективности поиска ближайшего соседа (1) более подробно.

2.2 Приближенный поиск ближайшего соседа

Один из наиболее перспективных способов повышения вычислительной эффективности полного перебора (1) связан с применением быстрых алгоритмов поиска ближайших соседей, которые можно разделить на две категории [4]. К первой категории относятся структуры для точного поиска ближайшего соседа, такие как k-d деревья, quadra-деревья и треугольные деревья. К сожалению, такой подход позволяет добиться прироста производительности лишь для векторов из 2–3 признаков, однако при работе с векторами признаками большой размерности ($M > 10$) их производительность становится существенно ниже производительности полного перебора за счет дополнительных накладных расходов на построение и обход дерева эталонов [8]. В результате наиболее применимой на практике считается вторая категория методов для *приближенного* поиска ближайшего соседа, которые в большинстве случаев возвращают ближайшего соседа, и достаточно хороший эталон (в смысле близости расстояния до него от распознаваемого объекта X по отношению к расстоянию между X и эталоном – ближайшим соседом) в остальных случаях [4, 8]. Например, может осуществляться поиск эталона, расстояние до которого от входного объекта не превышает некоторый порог [7]

$$\rho(X, X_v) < \rho_0. \quad (2)$$

Вначале обычно наугад выбирается один из эталонов $X_{r_1}, r_1 \in \{1, \dots, R\}$ и вычисляется рассогласование $\rho(X, X_{r_1})$.

Если, он отвечает требованию условия останова (3.1), процесс поиска оптимального решения на нем и завершается. В противном случае по определенному правилу, специфичному для каждого способа поиска приближенного ближайшего соседа, выбирается следующий эталон X_{r_2} и

процесс поиска повторяется до тех пор, пока не будет выполнено условие (2) или количество вычисленных расстояний не превысит заданный порог $E_{max} \leq R$ [8].

В настоящей работе предложен следующий алгоритм выбора эталона на основе универсального статистического подхода. Будем рассматривать задачу распознавания как проверку R статистических гипотез $W_r, r = 1, 2, \dots, R$ о распределении входного объекта (случайного вектора) X . Пусть на k -м этапе были проверены эталоны X_{r_1}, \dots, X_{r_k} . Используя информацию о расстояниях $\rho(X, X_{r_1}), \dots, \rho(X, X_{r_k})$, найдем наиболее вероятный ближайший (1) эталон $X_{r_{k+1}}$.

Предположим, что расстояния до всех проверенных эталонов статистически независимы. Будем искать следующий эталон

с помощью максимизации условной совместной плотности вероятности вычисленных на предыдущих шагах расстояниях:

$$r_{k+1} = \underset{v \in \{1, \dots, R\} - \{r_1, \dots, r_k\}}{\operatorname{argmax}} \prod_{i=1}^k f\left(\rho\left(X, X_{r_i}\right) \middle| W_v\right), \quad (3)$$

где $f\left(\rho\left(X, X_{r_i}\right) \middle| W_v\right)$ – условная плотность вероятности

рассогласования $\rho\left(X, X_{r_i}\right)$ при справедливости гипотезы

W_v . Для оценки этой плотности воспользуемся известным

эмпирическим фактом [2, 6] – если мера близости $\rho(X, X_r)$ определяется как среднее расстояние между соответствующими значениями признаков изображений X и X_r , то $\rho(X, X_r)$ при справедливости гипотезы W_v

распределено нормально $N(\rho_{v,r}; \sigma^2)$. Здесь

$\rho_{v,r} = \rho(X_v, X_r)$, а σ^2 – неизвестная дисперсия.

Эмпирические исследования показывают [6], что для большой размерности признакового пространства дисперсии расстояний между случайными векторами из одной генеральной совокупности оказываются приблизительно одинаковыми. Поэтому далее для простоты предположим, что дисперсия σ^2 не зависит от r . Тогда условную плотность в (3) можно записать как

$$f\left(\rho\left(X, X_{r_i}\right) \middle| W_v\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(\rho\left(X, X_{r_i}\right) - \rho_{v,r_i}\right)^2}{\sigma^2}\right] \quad (4)$$

Выполнив несложные преобразования, правило (3) окончательно запишется в следующем виде

$$r_{k+1} = \underset{\mu \in \{1, \dots, R\} - \{r_1, \dots, r_k\}}{\operatorname{argmin}} \sum_{i=1}^k \varphi_{\mu}(r_i). \quad (5)$$

где

$$\varphi_{\mu}(r_i) = \left(\rho\left(X, X_{r_i}\right) - \rho_{\mu,r_i}\right)^2. \quad (6)$$

Последнее выражение хорошо согласуется с известной эвристикой [7] – вероятность выбора μ -го эталона на следующем шаге тем выше, чем ближе между собой

рассогласования $\rho\left(X, X_{r_i}\right)$ и $\rho_{\mu,r}$.

К сожалению, для практической реализации предложенного подхода (5), (6) требуется заранее вычислить матрицу попарных расстояний между всеми эталонами, поэтому сложность алгоритма по затратам памяти запишется как $O(R^2)$. В результате практическая реализация такого алгоритма для достаточно больших объемов базы данных R оказывается невозможной. Поэтому далее воспользуемся известным в области приближенного ближайшего соседа

приемом [3] – наугад выбираются $p \ll R$ опорных эталонных изображений $\{X_{r_1}, \dots, X_{r_p}\}$ и на предварительном этапе вычисляются расстояния между всеми эталонами и выбранными опорными элементами. Следующие эталоны для проверки выбираются из условия максимизации совместной плотности вероятности расстояний аналогично (3), но при этом на каждом следующем шаге $k > p$ информация о вычисленных расстояниях $\rho(X, X_{r_{p+1}}), \dots, \rho(X, X_{r_{k-1}})$ учитываться не будет:

$$r_{k+1} = \underset{v \in \{1, \dots, R\} - \{r_1, \dots, r_k\}}{\operatorname{argmax}} \prod_{i=1}^p f\left(\rho(X, X_{r_i}) \Big| W_v\right), \quad (7)$$

Тогда предлагаемый алгоритм сведется к следующему. В процессе распознавания в начале вычисляются p рассогласований $\rho(X, X_{r_1}), \dots, \rho(X, X_{r_p})$. Далее для каждого оставшегося эталона $\mu \in \{1, \dots, R\} - \{r_1, \dots, r_p\}$ вычисляется следующая величина

$$\varphi_{\Sigma; \mu} = \sum_{i=1}^p \varphi_{\mu}(r_i). \quad (8)$$

Далее выполняется сортировка последовательности $\{\varphi_{\Sigma; \mu}\}$ и последовательно выбираются эталонные изображения с минимальным значением $\varphi_{\Sigma; \mu}$. Процесс поиска останавливается, если для очередного проверяемого эталона выполняется условие останова (2). Таким образом, для предлагаемого подхода (8) сложность по затратам памяти оказывается линейной $O(pR)$, что существенно ниже квадратичной сложности исходной реализации (5), (6). С учетом необходимости сортировки только наименьших E_{max} значений (partial sort), алгоритмическая сложность предложенной процедуры поиска приближенного ближайшего соседа в худшем случае оказывается равной $O(E_{max} \cdot M + R \log E_{max})$.

3. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

В настоящем разделе предложенный алгоритм (2), (6), (8) сопоставлялся с одними из лучших методов приближенного поиска ближайшего соседа – рандомизированным k-d деревом из библиотеки FLANN [8], основанным на выборе опорных эталонов методом Perm-sort [3] и быстрым поиском на основе структуры со свойством тесного мира (MSW, Metrized Small World) [4] из библиотеки NonMetricSpaceLib. Кроме того, был реализован предложенный для задачи статистической классификации метод максимально правдоподобного перебора (МПП) [9], основанный на асимптотических свойствах рассогласований между дискретными распределениями (расстояние хи-квадрат, дивергенции Кульбака-Лейблера и Йенсена-Шеннона и т.п.).

Для извлечения $M=4096$ признаков использовалась глубокая сверточная нейронная сеть, обученная для распознавания лиц в лаборатории Oxford Visual Geometry Group [5]. Для реализации применялась библиотека Caffe. Все значения

признаков нормировались так, чтобы их сумма была равна 1. В качестве мер близости (1) применялись расстояния Евклида и хи-квадрат. Для измерения среднего времени распознавания применялся ноутбук с процессором Intel Core i7-2630QM (тактовая частота 2 ГГц, объем ОЗУ 6 Гб).

В первом эксперименте использовался набор данных LFW (Labeled Faces in the Wild), являющейся стандартом де-факто для тестирования систем верификации и идентификации лиц (рис. 1).



Рис. 1: Фрагмент набора данных LFW

Были выбраны 9034 изображения $C = 1680$ людей, для которых доступно не менее двух фотографий. В обучающее множество включены $R = 4570$ фотографий, а тестирование проводилось на остальных 4464 изображениях. Оценки вероятности ошибки $\bar{\alpha}$ и среднего времени распознавания одного лица \bar{t} представлены в таблице 1.

Таблица 1. Результаты распознавания, LFW

| | Расстояние Евклида | | Расстояние хи-квадрат | |
|--|--------------------|----------------|-----------------------|----------------|
| | $\bar{\alpha}$, % | \bar{t} , мс | $\bar{\alpha}$, % | \bar{t} , мс |
| Полный перебор (1) | 15,21 | 12,85 | 14,83 | 65,85 |
| Рандомизированное k-d дерево | 16,33 | 4,82 | 16,22 | 7,39 |
| Perm-sort | 15,44 | 2,70 | 15,28 | 8,19 |
| MSW | 16,58 | 2,13 | 17,98 | 7,20 |
| МПП | 15,30 | 3,67 | 14,98 | 4,39 |
| Предложенный подход (2), (6), (8) | 15,17 | 1,37 | 14,98 | 3,91 |

В этом эксперименте вероятность ошибки для расстояния хи-квадрат на 0,4% ниже по сравнению с традиционной метрикой Евклида, однако время распознавания оказалось в 5 раз выше. Точность рандомизированного k-d дерева и MSW оказывается на 1-3% ниже точности полного перебора (1). В то же время предложенный подход наряду с perm-sort и МПП позволяют обеспечить вероятность ошибки распознавания, сопоставимую с методом ближайшего соседа (1). При этом метод Perm-sort позволил в 4,7–8 раз ускорить классификацию по сравнению с полным перебором. Наконец, предложенный в работе алгоритм при высокой точности характеризуется наилучшей вычислительной эффективностью: в 9–17 раз быстрее по сравнению с полным перебором и в 1,6–1,8 раз быстрее, чем традиционные методы приближенного ближайшего соседа. Отметим, что среднее

время классификации для нашего алгоритма оказалось сходным с методом МПП при минимизации расстояния хи-квадрат. Это неудивительно, т.к. метод МПП, как говорилось выше, специально разрабатывался с учетом асимптотических свойств подобных расстояний [9].

Во втором эксперименте использовался более простой набор данных PubFig83 из 13813 фотографий лиц (рис. 2). В обучающее множество были отобраны $R=5520$ изображений $C=83$ людей, а тестовое множество содержало остальные 8293 фотографии тех же людей. Результаты этого эксперимента приведены в таблице 2.



Рис. 2: Фрагмент набора данных PubFig83

Таблица 2. Результаты распознавания, PubFig83

| | Расстояние Евклида | | Расстояние хи-квадрат | |
|--|--------------------|----------------|-----------------------|----------------|
| | $\bar{\alpha}$, % | \bar{t} , мс | $\bar{\alpha}$, % | \bar{t} , мс |
| Полный перебор (1) | 6,36 | 15,28 | 6,18 | 82,97 |
| Рандомизированное k-d дерево | 6,83 | 2,25 | 6,31 | 5,22 |
| Perm-sort | 6,47 | 1,80 | 6,20 | 5,29 |
| MSW | 6,46 | 1,78 | 6,32 | 8,74 |
| МПП | 6,41 | 1,97 | 6,24 | 4,37 |
| Предложенный подход (2), (6), (8) | 6,46 | 1,21 | 6,31 | 2,66 |

Здесь результаты во многом аналогичны результатам для набора LFW (таблица 1), однако в связи с большим числом изображений для каждого класса в обучающей выборке точность метода ближайшего соседа (1) для различных мер близости оказывается практически одинаковой, при этом вероятность ошибки всех методов приближенного поиска также оказывается приемлемой. Предложенный подход позволил в 12-30 раз ускорить распознавание по сравнению с полным перебором. При этом разработанный алгоритм оказался в 1,5 раза эффективнее, чем наилучший из известных методов.

4. ЗАКЛЮЧЕНИЕ

В настоящей работе предложен новый алгоритм поиска приближенного ближайшего соседа в задаче распознавания лиц на основе статистического подхода (7) с использованием опорных элементов (8). Результаты экспериментов продемонстрировали повышенную вычислительную эффективность разработанного алгоритма по сравнению с известными методами. Дальнейшее развитие предложенного

подхода может заключаться в применении более точных оценок плотности вероятности в (3), (7), например, отказ от условия равенства дисперсий всех расстояний или замена нормального распределения на более подходящее распределений Вейбулла [2]. Кроме того, представляет интерес проведение экспериментальных исследования для больших баз данных, например, для набора Casia-WebFace.

6. СПИСОК ЛИТЕРАТУРЫ

- [1] Best-Rowden L., Han H., Otto C., Klare B.F., Jain A.K. Unconstrained face recognition: identifying a person of interest from a media collection // IEEE Transactions on Information Forensics and Security. 2014. Vol. 9. P. 2144-2157.
- [2] Burghouts G., Smeulders A., Geusebroek J.-M. The distribution family of similarity distances // Advances in Neural Information Processing Systems, 2008, P. 201-208.
- [3] Gonzalez E.C., Figueroa K., Navarro G. Effective proximity retrieval by ordering permutations // IEEE Transactions on PAMI 2008. Vol. 30, No. 9. P. 1647-1658.
- [4] Malkov Y., Ponomarenko A., Logvinov A., Krylov V. Approximate nearest neighbor algorithm based on navigable small world graphs // Information Systems. 2014. Vol. 45. P. 61-68
- [5] Parkhi O.M., Vedaldi A., Zisserman A. Deep face recognition // Proc. of the British Machine Vision. 2015. P. 6-17.
- [6] P'kalska E., Duin R.P. Classifiers for dissimilarity-based pattern recognition // Proc. of the 15th IEEE International Conference on Pattern Recognition, 2000. P. 12-16.
- [7] Savchenko A.V. Search Techniques in Intelligent Classification Systems. Switzerland: Springer International Publishing, 2016. 83 p.
- [8] Silpa-Anan C., Hartley R. Optimised KD-trees for fast image descriptor matching // Proc. IEEE Conf. on Computer Vision & Pattern Recognition, 2008. P. 1-8.
- [9] Савченко А.В. Метод максимально правдоподобного перебора в задаче классификации кусочно-однородных объектов // Автоматика и телемеханика. 2016. № 3. С. 99-108.