

ПАРАЛЛЕЛЬНОЕ РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ ЛИЦ МЕТОДОМ ГОЛОСОВАНИЯ С НЕЙРОСЕТЕВЫМ АРБИТРОМ

В.В. Пекунов

ООО «Спецлаборатория», Иваново

E-mail: pekunov@mail.ru

Задача классификации лиц, то есть идентификация личности по ее изображению, полученному, в частности, с камеры видеонаблюдения, весьма актуальна для современных систем безопасности. Среди методов идентификации выделяются: а) методы, основанные на выделении и сравнении характерных черт лица (метод эластичных графов и другие [6]), б) методы, основанные на редукции изображения всего лица к вектору проекций данного изображения на новый специфический базис (метод главных компонентов [3, 8] с последующим привлечением нейронных сетей, линейного дискриминантного анализа [3] и/или иных методов), в) методы, предполагающие какой-либо иной подход (скрытые марковские модели [2], дискретные преобразования Фурье) к анализу отдельных фрагментов изображения лица и их комплексов. Первый класс методов весьма чувствителен к точности определения координат черт лица и иногда требует наличия цветовой компоненты изображения [4], что в общем случае, когда изображение снимается с простых камер видеонаблюдения, часто черно-белых и не дающих качественного разрешения, затруднительно. Поэтому в сфере нашего особого внимания оказался второй класс методов, в основе которого лежит надежный и хорошо формализованный метод главных компонентов [8].

Задача данной работы состояла в том, чтобы построить надежный и быстродействующий классификатор лиц, представленных вектором координат в пространстве главных компонентов. Было выдвинуто требование работы всех компонентов системы в реальном масштабе времени, что определило потребность в максимально возможном применении технологий распараллеливания вычислений при решении данной задачи на отдельно взятом компьютере. Учитывая, что ресурсы центрального процессора обычно достаточно интенсивно задействованы под задачи видеонаблюдения и аналитики, весьма актуальным представляется использование вычислительных мощностей многоядерных видеокарт, многообразие которых закономерно приводит нас к идее применения кроссплатформенной технологии OpenCL.

Поступающие с камеры изображения лиц автоматически центрировались в результате работы детектора лиц (двухэтапной детекции с фильтром LBP (Local Binary Patterns, [7]) на первом этапе и отсеиванием результатов на втором этапе фильтром с примитивами Хаара [9]). Далее отсекался фон, для этого на изображении подбирался наиболее адекватный эллипс, выделяющий исключительно лицо, опираясь на его граничные точки, определенные детектором Кенни [1]. Подбор эллипса осуществлялся методом Хафа [1]. После отсека фона выполнялась эквализация изображения [1], позволяющая девальвировать различия в условиях освещенности.

После предварительной обработки к изображению лица применялся метод главных компонент, изображение проецировалось на базис из 45 собственных лиц, сгенерированных по базе BioID. Полученная метрика лица (вектор координат в пространстве главных компонентов) подлежала идентификации по базе данных метрик лиц персон. Сразу же было принято решение отказаться от чисто нейросетевой [5] идентификации с одной сетью-классификатором по причине ее высокой сложности и, следовательно,

чрезмерно высоких затрат на обучение и переобучение сети (при вводе в базу новых лиц), хотя и известно, что такой подход иногда дает весьма высокую вероятность распознавания. Однако справедливости ради следует отметить, что такой результат достигается при идентификации тех изображений лиц, которые непосредственно входят в обучающую выборку. При попытке идентификации лиц в несколько иных ракурсах (сдвинутых, наклоненных, повернутых) процент распознавания обычно уменьшается. Также мы отказались от одного из классических методов – линейного дискриминанта Фишера [3], поскольку сразу же оказалось, что ковариационная матрица часто не имеет полного ранга. Возможно помог бы отбор переменных (координат пространства главных компонентов) с применением корреляционного анализа, но было принято более простое решение – использовать для классификации линейный перцептрон.

В качестве возможных нейросетевых вариантов рассматривались лишь: а) *линейный перцептрон*¹, б) *классическая двуслойная нейронная сеть прямого распространения*² с первым нелинейным (экспоненциальные сигмоиды) слоем из 20÷32 нейронов и одним линейным нейроном во втором слое, обучение которой ведется градиентным методом (обратного распространения ошибки). Значительные надежды возлагались на *метод группового учета аргументов* (МГУА³, см. [3]) с простейшими квадратичными функциями-блоками, который позволяет инкрементально быстро строить не менее сложные и перспективные интерполяторы, чем нейронные сети. Кроме того, рассматривались такие элементарные методы как *наибольшей близости* [8] и *k-ближайших соседей* [3], базирующиеся на расстоянии Махалонобиса.

Было принято решение провести сравнительное экспериментальное исследование всех указанных методов. В трех первых случаях для каждой персоны строился отдельный интерполятор (сеть полиномов МГУА или нейронная сеть), который обучался выделять метрики лиц данной персоны среди других. Значение выхода интерполятора принадлежало диапазону [0; 1], величина значения определяла степень уверенности идентификации данной персоны. В таком случае при появлении в базе данных нового лица иногда можно оперативно переопределить (переобучить) требуемое количество интерполяторов за время, допустимое по требованиям решаемых нами задач.

Методы рассматривались на базе данных из 180 изображений лиц (28 персон в различных ракурсах). Задача была приближена к реальности: системе предъявлялись изображения персон в тех же ракурсах, что и в базе, а также и в несколько иных.

Таблица 1. Проценты верного распознавания персон различными методами

Метод				
Наибольшей близости	k-ближайших соседей	Линейные перцептроны	Двуслойные сети	МГУА
1. Изображения лиц в произвольных ракурсах				
77,5%	72,5%	67,5%	–	60%
2. Изображения лиц с принудительным варьированием ракурса				
83%	76,6%	55,3%	57,8%	55,3%

¹ Количество входов n равняется размеру вектора метрики лица. Сеть, в данном случае состоящая из одного нейрона со смещением, обучалась методом наименьших квадратов, сводящемуся к решению системы из $n+1$ линейных алгебраических уравнений.

² Количество входов равняется размеру вектора метрики лица. Все нейроны имели смещение.

³ В первом каскаде элементарных интерполяторов МГУА обычно осуществляется выбор из n^2 элементарных классификаторов, где n – число входов. В нашем случае для повышения скорости работы производился выбор из m^2 классификаторов, $m < n$, где m – количество наиболее «перспективных» входов, отобранных в ходе корреляционного анализа.

Примечательна *разница в качестве распознавания в первом и втором случаях*. Вероятно, именно лица в «обычных», наиболее часто встречающихся ракурсах, в предложенных условиях распознавались первыми двумя методами несколько хуже, чем лица в «существенно отклоняющихся» ракурсах. Для прочих методов ситуация обратная.

Лидируют два простых метода – наибольшей близости и k -ближайших соседей. Из трех интерполяционных методов наихудшие результаты показал МГУА, видимо потому, что полиномы дали достаточно маловероятное приближение, которое в сочетании с завышенной степенью таких полиномов привело к возникновению множества очагов неверной идентификации. Нейросетевые методы дали более правдоподобное приближение, но, видимо, проявился хорошо известный эффект [5] скорее чрезмерного подстраивания под элементы обучающей выборки чем выделения общей закономерности. Таким образом, для дальнейшей работы в качестве основных были выбраны оба вышеупомянутых простых метода в сочетании с принципом голосования. Персона считается идентифицированной, если за нее подали голос оба метода. Особенно же интересен случай «полемики» методов, когда они рекомендуют разные персоны.

При анализе экспериментальных данных по «полемике» было обнаружено, что помимо случаев, в которых оба простых метода ошибаются, есть ряд случаев, где хотя бы один из них дает верный ответ. Это позволило бы повысить точность предсказания, если бы удалось привлечь некоторый алгоритм-арбитр, который выберет верный ответ.

Достаточно сложной задачей оказался выбор арбитра. Было сделано предположение, что оказавшиеся более правдоподобными в сравнении с МГУА нейросетевые методы могут, если не идентифицировать одну персону среди множества, то, по меньшей мере, верно выбрать персону среди двух, предложенных методами наибольшей близости и k -ближайших соседей.

Были проведены дополнительные исследования, целью которых являлось определение наиболее перспективного метода-арбитра в случае «полемики» основных простых методов. Обнаружилось, что линейный персептрон дает верный ответ в 50% таких случаев, а двуслойная сеть – в 80% случаев. Таким образом, *предлагаемый в настоящей работе метод* идентификации является двухэтапным. На первом этапе проводится идентификация двумя первичными методами: а) поиск ближайшей (по Махаланобису) метрики в пространстве собственных лиц с рекомендацией соответствующей персоны; б) поиск k -ближайших метрик в том же пространстве с рекомендацией персоны, которой соответствует наибольшее количество этих метрик. Если рекомендации методов совпадают, то идентификация завершена. В противном случае начинается второй этап, на котором включается арбитр (либо линейные персептроны, либо двуслойные сети), выбирающий персону среди двух рекомендованных. Выбирается персона, для которой арбитр дает большее значение.

Требование быстрой работы определило следующий алгоритм работы арбитра:

а) если присутствуют обученные двуслойные нейронные сети для каждого из кандидатов, то выбирается персона, сеть которой дает лучший прогноз;

б) если обученных сетей нет, то либо они оперативно обучаются (если ресурсы вычислительной системы позволяют выполнить обучение за доли секунды) и работают в соответствии с пунктом а), либо задействуются линейные персептроны (если обученные персептроны отсутствуют, то они оперативно обучаются, это весьма быстрый процесс на современных ЭВМ).

В заключение рассмотрим подробно вопрос о максимально возможном повышении скорости обучения двуслойной нейронной сети. Методы, основанные на неполном переборе (например, случайного поиска) были отвергнуты по причине весьма значительных временных затрат на переобучение даже с применением параллельных вычислений. Метод обратного распространения ошибки (как и любой иной градиентный метод)

подразумевает *последовательную* (даже в случае какого-либо *распараллеливания* счета) *подстройку коэффициентов* для каждого элемента обучающей выборки размера k . Поэтому время обучения включает существенный компонент с прямой пропорциональностью k . Очевидно, что данный факт приводит к существованию разумного ограничения k_{max} на значение данной величины.

Предлагается включать в обучающую выборку лишь некоторые, наиболее характерные элементы. В нашем случае это можно сделать особенно просто, учитывая, что нейронная сеть является арбитром двух методов. В обучающую выборку *двухслойной сети*, сопоставленной каждой персоне, должны входить: а) лица, соответствующие данной персоне; б) лица, соответствующие «конкурирующим» персонам, то есть таким, с которыми у данной наиболее часто возникает «полемика» методов наибольшей близости и k -ближайших соседей; в) лица, соответствующие некоторому количеству случайным образом выбранных персон, что должно несколько повысить потенциальную надежность арбитража.

Для ускорения расчета было принято решение задействовать локальные вычислительные ресурсы многоядерных видеокарт. Этого должно было оказаться более чем достаточно для одновременного обучения двух и более нейронных сетей. Чтобы исключить привязку к конкретному производителю видеокарты, для распараллеливания был выбран интерфейс OpenCL.

Был выбран следующий *принцип распараллеливания*: каждая сеть отдавалась одному многопоточному мультипроцессору карты, каждый нейрон первого слоя сети обрабатывался отдельным потоком. Единственный нейрон второго, выходного слоя обрабатывался в каждом из этих потоков. Такая дублирующая схема оказалась наименее затратной, поскольку практически не требует координации потоков. При расчетах максимально возможно задействовалась внутренняя регистровая память потоков, там же, где это было невозможно, применялась разделяемая (shared в терминологии производителя видеокарт, она же local в терминологии OpenCL) память мультипроцессоров. При нехватке разделяемой памяти использовалась глобальная (преимущественно для передачи на видеокарту обучающей выборки).

Проводилось *сравнение результатов*, полученных с применением видеокарты, с результатами, полученными на классическом процессоре. Прямое сравнение временных характеристик обычной однопроцессорной программы и OpenCL-программы дало результат, на первый взгляд производящий впечатление абсурдного: расчет на 64 потоковых процессорах nVidia GeForce GTX570 оказался в 217 раз быстрее, чем на одном ядре процессора Intel Core i5-2300 (2,8 ГГц). Однако, если учесть, что однопроцессорная программа не была оптимизирована (Debug-версия компилятора C++ среды Microsoft Visual Studio) и содержала значительное количество операций индексного доступа к элементам массивов, от которых была свободна программа для видеокарты (за счет перехода от индекса массива к номеру потока), а также OpenCL-программа работала преимущественно с быстродействующей регистровой памятью, становится ясно, что вышеуказанный результат возможен.

Тем не менее, для контроля было принято решение о *дополнительном сравнении* с результатами, полученными той же OpenCL-программой, но запущенной уже не на видеокарте, а на вышеуказанном многоядерном центральном процессоре (ЦП) Intel с применением среды Intel OpenCL SDK 1.5. В данном случае расчет на видеокарте занял в 2,1 раза меньше времени чем расчет на четырех ядрах ЦП, то есть примерно в $2,1 \times 4 = 8,4$ раза меньше, чем расчет на одном ядре ЦП. Данную величину (8,4) можно было бы обозначить как «физическое ускорение» P_{opt} , помимо которого нас интересует также «логическое ускорение» S_{opt} по отношению к одному ядру ЦП, определяемое масштабированием результата с учетом разницы тактовых частот. Тактовая частота f_{cpu}

используемого ЦП (2,8 ГГц) в два раза выше тактовой частоты f_{gpu} потоковых (шейдерных) процессоров (1,4 ГГц) видеокарты. Тогда «логическое ускорение» составит

$$S_{opt} = P_{opt} \frac{f_{cpu}}{f_{gpu}} = 8,4 \frac{2,8}{1,4} = 16,8.$$

Эта величина приблизительно в 4 раза меньше ожидаемой (64-х, поскольку было задействовано именно такое количество потоковых процессоров). Вероятно, такой дополнительный множитель объясняется меньшей эффективностью исполнения серии команд потоковыми процессорами, то есть меньшим количеством команд, исполняемым за такт, в сравнении с ЦП Intel, что связано, в частности, с применением векторных SSE-команд ЦП.

По той же формуле легко подсчитать «логическое ускорение» S_{nopt} по сравнению с расчетом неоптимизированной программой, работающей на одном ядре ЦП:

$$S_{nopt} = 217 \frac{2,8}{1,4} = 434.$$

Резюмируя, в данной работе предлагается новый метод идентификации лиц, включающий опрос методов наибольшей близости и k -ближайших соседей с последующим использованием нейронной сети (двуслойной или линейной однослойной) в качестве арбитра в случае «полемики» методов. Показано, что применение арбитра позволяет повысить точность прогнозирования до 87÷88% при сохранении высокой скорости идентификации. Это достаточно хороший показатель, особенно с учетом того, что идентифицирующей системе предъявлялись изображения лиц не только из базы данных, но схожие с ними, в том числе повернутые и наклоненные. При обучении двуслойных нейронных сетей применен перенос вычислений (с распараллеливанием) на видеокарту. На 64 потоковых процессорах nVidia GeForce GTX570 получено существенное логическое ускорение в 16,8 раз (физическое – в 8,4 раза) *по сравнению с оптимизированным расчетом* на одном ядре процессора Intel Core i5-2300 (2,8 ГГц) и в 434 раза (физическое – в 217 раз) *по сравнению с неоптимизированным расчетом* на том же процессоре.

Литература

1. Визильтер Ю.В. Обработка и анализ цифровых изображений с примерами на LabVIEW и IMAQ Vision / Ю.В. Визильтер, С.Ю. Желтов, В.А. Князь, А.Н. Ходарев, А.В.Моржин. – М.: Книга по требованию, 2011. – 464 с.
2. Гультяева Т.А., Попов А.А. Скрытые марковские модели с одномерной топологией в задаче распознавания лиц // Сборник научных трудов НГТУ. – 2006. – № 1. – С. 1–6.
3. Дюк В., Самойленко А. Data mining: учебный курс – СПб: Питер, 2001. – 368 с.
4. Лопатина А.Д. Выделение области лица с помощью комбинации методов цветовой и яркостной сегментации // Вестник УГАТУ. Управление, ВТ и И. – 2009. – Т.13, №2 (35). – С.180–187.
5. Рассел С., Норвиг П. Искусственный интеллект: современный подход – М.: «Вильямс», 2007. – 1408 с.
6. Самаль Д.И. Алгоритмы идентификации человека по фотопортрету на основе геометрических преобразований: автореф. дис. ... канд. тех. наук – Минск, 2002. – 20 с.
7. Ahonen T., Hadid A., Pietikäinen M. Face recognition with local binary patterns // Proc. Eighth European Conf. Computer Vision – Springer, 2004. – P. 469–481.
8. Turk M., Pentland A. Eigenfaces for recognition // Journal of Cognitive Neuroscience. – 1991. – Vol. 3. – №1. – P. 71–86.
9. Viola P.A., Jones M.J. Rapid Object Detection using a Boosted Cascade of Simple Features // Computer Vision Pattern Recognition. 2001. (1). P. 511–518.