

АНДРЕЙ НИКОЛАЕВИЧ ТАЛБОНЕН
преподаватель кафедры теории вероятностей и анализа данных математического факультета, Петрозаводский государственный университет (Петрозаводск, Российская Федерация)
antal@sampo.ru

АЛЕКСАНДР АЛЕКСАНДРОВИЧ РОГОВ
доктор технических наук, профессор, заведующий кафедрой теории вероятностей и анализа данных математического факультета, Петрозаводский государственный университет (Петрозаводск, Российская Федерация)
rogov@psu.karelia.ru

МОДЕЛИ И МЕТОДЫ ПОИСКА ЛЮДЕЙ НА ФОТОГРАФИЯХ ИЗ ИСТОРИЧЕСКОГО АЛЬБОМА*

Описываются методы детектирования и распознавания лиц, основанные на локальных бинарных шаблонах. Предложенные методы ориентированы на изображения невысокого качества, характерные для коллекций исторических фотографий.

Ключевые слова: детектирование лиц, распознавание лиц, алгоритм Виолы – Джонса, локальные бинарные шаблоны

ВВЕДЕНИЕ

На данный момент существует множество работ по распознаванию лиц и разнообразное готовое программное обеспечение, позволяющее обнаруживать лица и сравнивать их между собой. Как правило, детектирование и распознавание лиц выполняется разными алгоритмами. В обоих случаях качество работы алгоритма, выраженное количественными характеристиками точности и полноты [1], зависит от качества исходных данных. В большинстве публикаций описываются результаты работы алгоритмов на качественных изображениях [2], [3], [4]. Отсутствуют работы по исследованию качества работы алгоритмов на изображениях низкого качества. Однако существуют коллекции с фотографиями низкого качества, на которых тоже требуется осуществлять поиск. К ним относятся цифровые исторические альбомы.

Исходными данными для проведенных исследований является электронная коллекция черно-белых изображений строительства Беломорско-Балтийского канала (ББК) в формате JPEG, созданных около 80 лет назад. Коллекция насчитывает более 6 тыс. изображений. Преобладание диапазона лиц размером 20–40 пикселей является следствием низкого качества как оригинальных фотографий, так и оцифрованных изображений.

Для алгоритмов детектирования характерны ложные срабатывания, когда определенный участок изображения, не являющийся лицом, но содержащий некоторые локальные признаки, признается классификатором как лицо. Вследствие этого снижается точность результатов. При этом наличие большого количества деталей и шумов повышает вероятность ложного сраба-

тывания. Изменяя параметры алгоритма, можно оказывать влияние на вероятность обнаружения лица или ложного срабатывания. Но при этом в рамках одного алгоритма можно повысить точность за счет полноты и наоборот. Таким образом, можно наблюдать некоторую обратную зависимость, не позволяющую повысить оба показателя. Для решения этой проблемы можно использовать дополнительную классификацию: сортировку результатов детектирования с наибольшей полнотой.

ОСОБЕННОСТИ ЗАДАЧИ

В данной работе для обнаружения лиц используется алгоритм Виолы – Джонса [10]. Данный алгоритм был выбран потому, что он обладает высокими показателями по сравнению с другими бесплатными / широкодоступными алгоритмами [5], [11]. В процессе исследования с помощью этого алгоритма были получены коллекции с наибольшей полнотой («полная» коллекция) и с наибольшей точностью («точная» коллекция). Одной из задач данной работы является построение классификатора, позволяющего повысить точность «полной» коллекции, обеспечить показатели отсортированной коллекции выше показателей «точной».

Для распознавания лиц широко используется алгоритм локальных бинарных шаблонов (ЛБШ). В некоторых исследованиях алгоритм ЛБШ работал с более высокой точностью, чем другие алгоритмы [2].

РАСПОЗНАВАНИЕ ОБЪЕКТОВ С ПОМОЩЬЮ LBP

LBP, впервые описанный в [8], представляет собой простой и мощный инструмент для распознавания различных элементов изображения.

В качестве пространства признаков используются гистограммы так называемых кодов LBP. Благодаря высокой скорости расчета некоторых типов кодов LBP стал широко применяться для распознавания образов.

LBP представляет собой фильтр, обозначаемый как $LBP_{P,R}(x, y)$, который для каждой точки изображения рассчитывает код на основе значений точек в некоторой окрестности этой точки. В данном случае P – число точек, R – радиус окрестности. Точки окрестности обозначим как g_i , где $i = \overline{0, P-1}$. При этом координаты точки рассчитываются как $(R \cdot \cos(\frac{i}{2\pi}); R \cdot \sin(\frac{i}{2\pi}))$. Обозначим изображение как $f(x, y)$ или как $f(g)$, если g – точка.

Пусть $s_i(x, y) = \begin{cases} 1, f(g_i) > f(g_c) \\ 0, \text{иначе} \end{cases}$, где g_c – точка с координатами (x, y) . Тогда $LBP_{P,R}(x, y) = \sum_{i=0}^{P-1} 2^i \cdot s_i(x, y)$.

Таким образом, последовательность $s_i(x, y)$, где $i = \overline{0, P-1}$, представляет собой двоичную последовательность кода LBP. Следовательно, $LBP_{P,R}(x, y) \in [0, 2^P - 1]$.

Для сравнения двух изображений в качестве векторов признаков используются гистограммы кодов LBP. В общем случае для каждого изображения строится гистограмма $H(l)$ для значений $LBP_{P,R}(x, y)$, где $l = \overline{0, P-1}$. Существует несколько методов расчета расстояния между гистограммами. Например, расстояние Хи-квадрат:

$$\chi^2(H_1, H_2) = \sum_{i=0}^{B-1} \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)}, \text{ где } B \text{ – число кодов.}$$

Существуют также модификации фильтра, описанные в работах [6], [7], [8], [9]:

1. Некоторые коды несут в себе больше информации, чем другие. Коды, в двоичной циклической записи которых число переходов между последовательностями «1» и «0» не превышает двух, обозначаются как «uniform», что соответствует слову «равномерный» [3]. Для заданного P существует всего $P \cdot (P - 1) + 2$ равномерных значений. Модифицированный фильтр $LBP_{P,R}^{u2}$ в этом случае возвращает коды равномерных значений, добавляя только один код для неравномерных значений.

$$LBP_{P,R}^{u2}(x, y) = \begin{cases} \text{индекс кода, если он равномерный} \\ P \cdot (P - 1) + 2, \text{иначе} \end{cases}$$

2. Так как окрестность представляет собой круг, можно найти группы кодов, инвариантных к повороту. Для каждого кода LBP существует P кодов, инвариантных к повороту, получаемых путем циклического сдвига P -битового числа. Для каждой такой группы в фильтр попадает минимальное значение кодов данной группы. Задача определения количества кодов, инвариантных к повороту, является нетривиальной.

Фильтр обозначается как $LBP_{P,R}^n$.

3. С учетом предыдущих двух свойств определяется также равномерный фильтр, инвариантный к повороту. Кодов LBP, обладающих одновременно двумя свойствами, всего $P + 2$, которые отличаются друг от друга числом бит, равных 1. В этом случае фильтр $LBP_{P,R}^{ru2}$ задается следующим образом:

$$LBP_{P,R}^{ru2}(x, y) = \begin{cases} \text{число единиц, если код равномерный} \\ P + 1, \text{иначе} \end{cases}$$

В работе [8] был предложен метод сравнения вычисления гистограмм лиц. Изображение лица разбивается на $k \times k$ участков, для каждого из которых рассчитывается гистограмма. Итоговая гистограмма изображения лица определяется как конкатенация гистограмм участков изображения.

В работе [4] был предложен расширенный метод сравнения гистограмм лиц, основанный на взвешенной матрице. задается матрица весов $k \times k$, каждый элемент которой соответствует участку изображения. Обозначим гистограмму j -го участка изображения как H^j , $j = \overline{0, k^2 - 1}$. Для каждого j -го участка задается вес w_j . Тогда можно определить модифицированное (взвешенное) расстояние Хи-квадрат следующим образом:

$$\chi_w^2(H_1, H_2) = \sum_{j=0}^{k^2-1} w_j \cdot \sum_{i=0}^{B-1} \frac{(H_1^j(i) - H_2^j(i))^2}{H_1^j(i) + H_2^j(i)},$$

где B – число паттернов.

В частности, в [4] предлагается использовать следующую матрицу для распознавания лиц (рис. 1).

2	1	1	1	1	1	2
2	4	4	3	4	4	2
1	1	1	2	1	1	1
0	1	1	2	1	1	0
0	1	1	1	1	1	0
0	1	1	2	1	1	0
0	1	1	1	1	1	0

Рис. 1. Матрица весов для распознавания лиц

ПОВЫШЕНИЕ ПОЛНОТЫ И ТОЧНОСТИ ДЕТЕКТИРОВАНИЯ ЛИЦ

Недостатком «полной» коллекции является большое число ложных объектов. Путем отсеивания найденных объектов с помощью обучающего множества можно повысить точность результатов.

Задача классификации сводится к выбору метода сравнения двух объектов и формированию множеств лиц и ложных объектов, которые будут являться обучающей выборкой.

Пусть F^* – множество лиц, а E^* – множество ложных объектов из обучающей выборки. Пусть $\rho_f(f_i, f_j)$ – расстояние между объектами. Суть алгоритма классификации заключается в сравнении расстояния между проверяемым объектом и множеством F^* и расстояния от объекта до E^* . Если проверяемый объект ближе к множеству F^* , то объект считается лицом, иначе – ложным объектом.

Поскольку любая классификация предполагает ненулевую вероятность ошибок 1-го и 2-го рода, алгоритм вычисления расстояния и обучающие множества подбираются для уменьшения количества ошибок обоих типов.

Применение данного алгоритма классификации над множеством с «наибольшей полнотой» («полная» коллекция, множество является результатом работы алгоритма детектирования лиц с наибольшей полнотой) позволяет за счет определенного количества ошибок 1-го рода (лица, признанные ложными объектами) повысить точность для данного множества.

ЭКСПЕРИМЕНТЫ

В процессе исследования коллекции было обнаружено около 4 тыс. лиц. Для оценки качества найденных объектов по их размерам было составлено общее распределение (рис. 2).

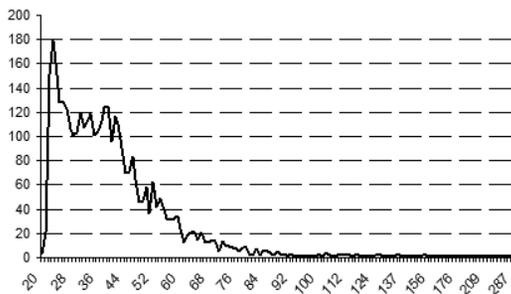


Рис. 2. Распределение размеров «полной» коллекции

Для оценки полноты и точности необходимо произвести экспертную оценку тестовой выборки (экспертная коллекция). В экспериментах использовалась выборка примерно в 1070 изображений. Процесс оценивания был упрощен: вместо выделения всех лиц на изображениях автоматически выделялись объекты полной коллекции, отсортированной вручную, после чего эксперту достаточно было отметить на изображении недостающие лица. Ручная сортировка также осуществлялась быстрым методом, позволяющим сортировать 100 объектов за 1 минуту.

Кроме очевидных лиц и ложных объектов, найденных в «полной» коллекции, были определены дополнительные правила оценивания. К ложным объектам относились слабо различимые лица (не видно глаз или рта) и лица людей, повернувшихся к камере на угол больше 90 градусов. К лицам относились различные изображения лиц: портреты, рисунки, бюсты.

В процессе классификации объекты преобразовывались к размеру 64×64 . Эксперименты проводились и с использованием нескольких фильтров с различными параметрами расчета гистограммы и с различными обучающими множествами. Для вычисления векторов признаков использовались фильтры $LBP_{16,3}^{riu2}$ и $LBP_{24,3}^{riu2}$. В одних случаях применялась матрица весов (рис. 1), в других случаях вектор признаков сжимался до размера 200. В качестве множества лиц использовался набор из 18 изображений лиц. Кроме того, были заданы 2 множества ложных объектов (8 и 26 изображений, E_1 и E_2 соответственно).

Результатом каждого эксперимента являлась отдельная коллекция объектов. Полученные коллекции оценивались полнотой и точностью, для расчета которых объекты коллекций сопоставлялись с объектами экспертной коллекции. Результаты экспериментов, позволяющие сравнить качество полученных коллекций с соответствующими параметрами «полной» и «точной» коллекций, представлены в табл. 1 и на рис. 3. Описание каждого эксперимента содержит параметры фильтра LBP, использованные методы расчета гистограмм, а также множество ложных объектов.

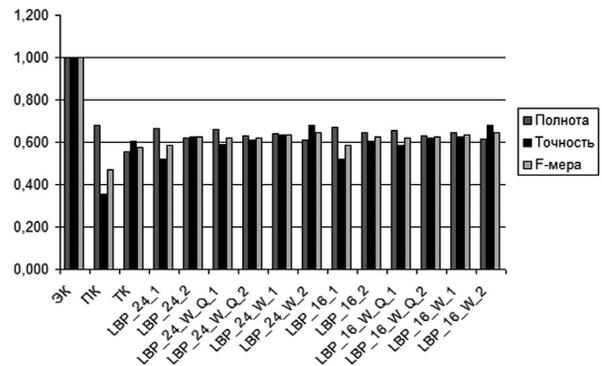


Рис. 3. Диаграмма сравнения полноты и точности обнаружения лиц

Проведенные эксперименты показали, что гистограммы, полученные с помощью фильтров $LBP_{16,3}^{riu2}$ и $LBP_{24,3}^{riu2}$ матрицы весов и расширенной обучающей выборки, обладают наилучшей селективной способностью с точки зрения F-меры [1]. Кроме того, в большинстве случаев расширение обучающей выборки позволило сократить количество ошибок и, следовательно, точность и F-меру отсортированной коллекции.

РАСПОЗНАВАНИЕ ЛИЦ С ПОМОЩЬЮ LBP

Современные исследования демонстрируют высокие показатели точности распознавания лиц размерами выше 100. В данном случае требуется отдельный подход. Для экспериментов использовалась следующая коллекция изображений (табл. 2), обозначим данный набор как T .

Таблица 1

Результаты экспериментов по повышению точности и полноты

Название	Описание	Лица	Ошибки	Полнота	Точность	F-мера
ЭК	Экспертная коллекция	793	0	1,000	1,000	1,000
ПК	Полная коллекция	540	973	0,681	0,357	0,468
ТК	Точная коллекция	439	289	0,554	0,603	0,577
LBP_24_1	$LBP_{24,3}^{riu2}, E_1$	529	484	0,667	0,522	0,586
LBP_24_2	$LBP_{24,3}^{riu2}, E_2$	492	294	0,620	0,626	0,623
LBP_24_W_Q_1	$LBP_{24,3}^{riu2}$, веса, сжатие, E_1	522	364	0,658	0,589	0,622
LBP_24_W_Q_2	$LBP_{24,3}^{riu2}$, веса, сжатие E_2	499	319	0,629	0,610	0,619
LBP_24_W_1	$LBP_{24,3}^{riu2}$, веса, E_1	506	292	0,638	0,634	0,636
LBP_24_W_2	$LBP_{24,3}^{riu2}$, веса E_2	483	226	0,609	0,681	0,643
LBP_16_1	$LBP_{16,3}^{riu2}, E_1$	532	492	0,671	0,520	0,586
LBP_16_2	$LBP_{16,3}^{riu2}, E_2$	512	332	0,646	0,607	0,626
LBP_16_W_Q_1	$LBP_{16,3}^{riu2}$, веса, сжатие, E_1	520	370	0,656	0,584	0,618
LBP_16_W_Q_2	$LBP_{16,3}^{riu2}$, веса, сжатие, E_2	499	306	0,629	0,620	0,625
LBP_16_W_1	$LBP_{16,3}^{riu2}$, веса, E_1	510	305	0,643	0,626	0,634
LBP_16_W_2	$LBP_{16,3}^{riu2}$, веса, E_2	488	230	0,615	0,680	0,646

Таблица 2

Тестовое множество изображений для распознавания лиц



Были обнаружены следующие пары лиц, соответствующие одному и тому же человеку: {3,

14}, {4, 13} и {7, 9}. Обозначим данное множество как T^+ . Кроме того, очевидна пара ложных объектов, которые должны быть близки друг к другу: {1, 15} (T^-). Дополним множества T^+ и T^- зеркально отраженными элементами данных множеств.

Для исследования работы различных фильтров LBP используется следующий алгоритм:

1. Для каждого объекта $t_i \in T$ находится ближайший к нему объект $t_i^* \in T$. То есть

$$t_i^* = \arg \min_{t_j} \{\rho_f(t_i, t_j) \mid t_j \in T / t_i\}.$$

2. Полученное таким образом множество пар $T^P = \{ \langle t_i, t_i^* \rangle \mid t_i \in T \}$ сопоставляется с известными парами лиц и ложных объектов.

3. Обозначим $h(t_i, t_j) = \begin{cases} 1, & \langle t_i, t_j \rangle \in T^+ \text{ или } \langle t_i, t_j \rangle \in T^- \\ 0, & \text{иначе} \end{cases}$.

4. Будем рассчитывать точность следующим образом:

$$\text{Re} = \frac{\sum_{t_i \in T} h(t_i, t_i^*)}{|T^+| + |T^-|}.$$

В процессе исследования были рассмотрены различные варианты фильтров LBP. В табл. 3 приведено сравнение точности использованных фильтров относительно заданного множества T .

Таблица 3

Результаты экспериментов по распознаванию лиц

Обозначение	Точность Re	Описание
LBP _{8,1}	3/8	P = 8, R = 1
LBP _{16,1}	2/8	P = 16, R = 1
LBP _{8,2}	4/8	P = 8, R = 2
LBP _{16,2}	4/8	P = 16, R = 2
LBP _{8,3}	4/8	P = 16, R = 3
LBP _{16,3}	6/8	P = 16, R = 3
Взвешенный LBP _{16,3} ⁿ	4/8	P = 16, R = 3, инвариантный
Взвешенный LBP _{16,3} ^{nu}	3/8	P = 16, R = 3, равномерный, инвариантный
Взвешенный LBP _{16,3} ^u	5/8	P = 16, R = 3, равномерный
Взвешенный LBP _{16,3}	8/8	P = 16, R = 3

Таким образом, наибольшая точность достигается для взвешенного фильтра LBP_{16,3}.

АННОТИРОВАНИЕ ЛИЦ

Аннотирование изображений предполагает присвоение каждому изображению набора текстовых меток, соответствующих данному изображению. При аннотировании за счет лиц необходима база данных лиц. Имена людей из базы данных будут являться текстовыми метками. При наличии такой базы данных предлагается

использовать взвешенный фильтр LBP_{16,3} для поиска похожих лиц.

В случае если база данных отсутствует, исходя из результатов экспериментов предлагается следующая схема аннотирования изображений с помощью эксперта:

1. Для каждого найденного лица выполняется поиск нескольких ближайших лиц, вычисленные расстояния между объектами сохраняются в базе данных. Количество ближайших объектов при этом задается.
2. При оценивании каждого лица эксперту предлагается сравнить его с ближайшим.
3. Если эксперт принимает предложенный объект, система запоминает новую связь.
4. Если эксперт отклоняет предложенный объект, система предлагает ему следующий ближайший объект по списку.

ЗАКЛЮЧЕНИЕ

Были предложены методы повышения полноты и точности алгоритма Виолы – Джонса с помощью локальных бинарных шаблонов. Для распознавания лиц низкого качества была найдена модификация LBP, позволяющая сравнивать лица с максимальной точностью.

Для случая отсутствия базы данных лиц был предложен метод аннотирования изображений с помощью эксперта на основе локальных бинарных шаблонов.

* Работа выполнена при финансовой поддержке Программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

СПИСОК ЛИТЕРАТУРЫ

1. Агеев М., Кураленок И. Официальные метрики РОМИП'2004 [Электронный ресурс]. Режим доступа: http://romip.ru/docs/romip_metrics.pdf
2. Маслий Р. В. Использование локальных бинарных шаблонов для распознавания лиц на полутоновых изображениях [Электронный ресурс]. Режим доступа: http://www.nbu.gov.ua/e-journals/vntu/2008-4/2008-4_ru.files/ru/08rvmgisi_ru.pdf
3. Петрук В. И., Самородов А. В., Спиридонов И. Н. Применение локальных бинарных шаблонов к решению задачи распознавания лиц // Вестник МГТУ им. Н. Э. Баумана. Сер. «Приборостроение». 2011. Спец. вып. Биометрические технологии. С. 58–63.
4. Ahonen T., Hadid A., Pietikäinen M. Face Recognition with Local Binary Patterns [Electronic resource]. Access mode: <http://masters.donntu.edu.ua/2011/ft/dyru/library/article8.pdf>
5. Degtyarev N., Seredin O. Comparative Testing of Face Detection Algorithms [Electronic resource]. Access mode: <http://lda.tsu.tula.ru/papers/degyarev-2010-icisp-ctfd.pdf>
6. Guo Z., Zhang L., Zhang D. Rotation invariant texture classification using LBP variance (LBPV) with global matching [Electronic resource]. Access mode: http://www4.comp.polyu.edu.hk/~cslzhang/paper/PR_10_Mar_LBPV.pdf
7. Mäenpää T. The local binary pattern approach to texture analysis – extensions and applications [Electronic resource]. Access mode: <http://herkules.oulu.fi/isbn9514270762/isbn9514270762.pdf>
8. Ojala T., Pietikäinen M., Harwood D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions // Pattern Recognition. 1996. Vol. 29. № 1. С. 51–59.
9. Ojala T., Pietikäinen M., Mäenpää T. A Generalized Local Binary Pattern Operator for Multiresolution Gray Scale and Rotation Invariant Texture Classification [Electronic resource]. Access mode: <http://www.mediateam.oulu.fi/publications/pdf/43.pdf>
10. Viola P., Jones M. Robust Real-time Object Detection [Electronic resource]. Access mode: http://research.microsoft.com/en-us/um/people/viola/Pubs/Detect/violaJones_IJCV.pdf
11. Wechsler H. Reliable face recognition methods: system design, implementation and evaluation [Electronic resource]. Access mode: http://books.google.ru/books?id=refsB92dvEC&printsec=frontcover&hl=ru&source=gbs_atb#v=onepage&q&f=false