

И.В. Бекетова, С.Л. Каратеев, Ю.В. Визильтер, канд. техн. наук, А.В.Бондаренко, канд. ф.-м. наук, С.Ю. Желтов, д-р техн. наук (ГосНИИАС)

Автоматическое обнаружение лиц на цифровых изображениях на основе метода адаптивной классификации *AdaBoost*

Описан метод автоматического обнаружения и локализации изображений лиц на цифровых изображениях. При построении алгоритмов обнаружения использовался модифицированный алгоритм синтеза высокоэффективных классификаторов из множества слабых классификаторов типа Adaboost, отличающийся от классического варианта использованием аппроксимации распределения вероятностей откликов слабых классификаторов вместо пороговой решающей функции.

Keywords: Face detection, classification, boosting.

Работа выполнена при поддержке РФФИ, проект 05-08-18234-а.

Введение

В настоящее время, в связи с широким распространением автоматизированных систем видеонаблюдения и систем биометрической идентификации личности особое практическое значение приобретает проблема автоматического обнаружения и локализации человеческих лиц на изображениях, получаемых от камер видеонаблюдения, а также различных систем контроля доступа, паспортного контроля и т.д.

Задача обнаружения изображений лиц представляет собой задачу поиска и распознавания на цифровом изображении локальных областей, обладающих специфическими характерными для изображений лиц параметрами. Поскольку, как правило, в таких системах отсутствует априорная информация о количестве, размерах и местоположении изображений лиц на цифровом изображении, для решения задачи оказывается необходимо проанализировать каждый фрагмент изображения, выделенный скользящим окном разных масштабов. При этом результатом анализа фрагмента изображения является классификация фрагмента – принятие решения о принадлежности его к классу изображений лиц или классу фоновых изображений. Таким образом, задачу обнаружения и локализации изображений лиц можно рассмотреть как частный случай задачи классификации.

Как правило, решение задачи классификации имеет вид следующей последовательности действий:

- формирование признакового пространства;
- построение системы принятия решения в признаковом пространстве;
- классификация предъявляемого объекта путем проверки системой принятия решения гипотезы о принадлежности данного объекта к классу искомым объектов.

Формирование признакового пространства и построение системы принятия решения в признаковом пространстве осуществляется путём предварительного обучения классификатора на выборке эталонных изображений.

Для каждой конкретной задачи классификатор может быть построен на базе различных методов и подходов, таких как байесовские решающие правила, искусственные нейронные сети, метод опорных векторов, метод усиления слабых классификаторов и т.д. В каждом конкретном случае метод классификации выбирается, исходя из требований эффективности классификации по таким показателям как точность, устойчивость к искажениям, инвариантность к условиям наблюдения, вычислительная производительность, минимальные затраты на обучение и тестирование. К сожалению, для большинства подходов эти требования оказываются противоречивыми.

В последнее десятилетие появился метод решения задачи классификации, удачно сочетающий высокое качество классификации с высокой вычислительной производительностью – метод усиления (boosting) слабых (weak) классификаторов [1]. Метод основан на комбинировании «слабых» классификаторов, точность которых может лишь незначительно превышать вероятность случайного угадывания (т.е. вероятность правильной классификации для каждого слабого классификатора немногим больше $\frac{1}{2}$). В результате объединения таких «слабых» классификаторов строится один «сильный» (strong), высокоэффективный классификатор. В настоящее время данный метод активно используется в различных прикладных задачах анализа изображений.

В статье представлено оригинальное приложение метода усиления слабых классификаторов к задаче обнаружения лиц на цифровых изображениях.

Метод адаптивного усиления слабых классификаторов *AdaBoost*

Базовый алгоритм адаптивного усиления *AdaBoost* был опубликован в статье Freund и Schapire в 1997 г. [1]. Основная идея алгоритма состоит в том, что в процессе обучения на обучающей выборке эталонных изображений (образцов) формируется последовательность слабых классификаторов. Откликом слабого классификатора является

классифицирующая гипотеза на пространстве входных признаков. Окончательная гипотеза формируется взвешенным голосованием гипотез слабых классификаторов. Обучение произвольного количества слабых классификаторов проводится на одной обучающей выборке, составленной из образцов, которым приписаны соответствующие весовые коэффициенты. На каждой итерации взвешивание образцов производится адаптивно по отношению к результатам работы предыдущего слабого классификатора: больший вес получают те примеры, на которых допускались ошибки предыдущими классификаторами.

Обобщённо алгоритм *AdaBoost* может быть представлен следующим образом [2].

1. Сформировать обучающую выборку из $i = 1, \dots, m$ эталонных изображений (x_i, y_i) с набором признаков $x_i \in \{x_1, \dots, x_k, \dots, x_n\}$ и метками классов $y \in \{1, -1\}$. Установить на начальном распределении D_1 одинаковые веса образцов

$$D_1 = 1/m.$$

Сумма всех весов, таким образом, приведена к единице.

2. Далее на каждой итерации $t = 1, \dots, T$ выполнять:

2.1. Выбрать на текущем распределении D_t слабый классификатор с наименьшей ошибкой ε_t классифицирующей гипотезы h_t

$$\varepsilon_t = \arg \min \{ P_{i \sim D_t} [y_i \neq h_t(x_i)] \}.$$

2.2. Вычислить коэффициент

$$\alpha_t = 1/2 \ln ((1-\varepsilon_t) / \varepsilon_t).$$

2.3. Обновить взвешенное нормированное распределение образцов

$$D_{t+1}(i) = (D_t(i) \exp[-\alpha_t y_i h_t(x_i)]) / Z_t,$$

$$\text{где } Z_t = \sum_i D_t(i) \exp[-\alpha_t y_i h_t(x_i)].$$

3. Сформировать финальную гипотезу-классификатор как линейную комбинацию классифицирующих гипотез слабых классификаторов

$$H(x) = \text{sign} \left\{ \sum_t \alpha_t h_t(x) / \sum_t \alpha_t \right\}$$

Из вышеприведённого следует, что на каждой итерации алгоритм адаптивно настраивается на ошибки предыдущих слабых классификаторов путём повышения относительного веса образцов, на которых допускались ошибки: наибольшие веса присваиваются образцам (x, y) с наименьшим значением

$$y \sum_{t'} \alpha_{t'} h_{t'}(x), \text{ где } t' = 1, \dots, t-1.$$

Эту величину можно интерпретировать как зазор или отступ (*margin*) между классами для конкретного образца (x, y) . Зазор положителен, когда образец классифицирован правильно, в противном случае зазор отрицателен. Алгоритм старается максимизировать зазор между классами путём последовательной минимизации на каждой итерации значения

$$Z_t = \sum_i D_t(i) \exp[-\alpha_t y_i h_t(x_i)]$$

как функции α_t .

Freund и Schapire доказали [3], что если ошибки слабых классификаторов ε_t меньше $1/2$ на всех распределениях взвешенных образцов D_t , так что $\varepsilon_t \leq 1/2 - \gamma_t$ и $\gamma_t > 0$, то ошибка обучения финального классификатора экспоненциально стремится к нулю по мере увеличения количества слабых классификаторов (числа итераций).

На каждой итерации обучения алгоритм работает с обучающей выборкой одного и того же фиксированного размера, поэтому производительность процесса обучения в смысле затрат вычислительных ресурсов определяется вычислительной эффективностью обучения слабых классификаторов. Целесообразно выбирать слабые классификаторы простой структуры. Достаточно лишь, чтобы ошибка обучения слабого классификатора не превышала ошибку угадывания

$$P_{i \sim D_t} [y_i \neq h_t(x_i)] \leq 1/2 - \gamma_t, \text{ где } \gamma_t > 0.$$

Часто в качестве слабого классификатора используется классификатор порогового типа, выдающий на выходе метку класса со значениями $[1, -1]$:

$$h_t(x_i = \{x_1, \dots, x_k, \dots, x_n\}) = \text{sign}(x_k - \theta_k),$$

где θ_k – порог для k -ой составляющей набора признаков x_i .

Пороговый классификатор предполагает унимодальность распределения для каждой составляющей x_k набора признаков x_i . Преодолеть данное ограничение можно путем использования аппроксимации распределения вероятностей откликов слабых классификаторов. При этом плотность вероятности оценивается статистически путем построения гистограммы распределения составляющих x_k , взвешенных по всей совокупности обучающей выборки $(1, \dots, m)$:

$$h_t(x_i) = \text{sign} \left\{ 1/2 \log \left(\frac{D_{y=+1} + \varepsilon}{D_{y=-1} + \varepsilon} \right) \right\}, \text{ где}$$

$D_{+1} = P[f(x_k) \in \text{bin}_j, y=+1]$ – степень достоверности того, что приведённое к диапазону $[0..1]$ значение составляющей признака $f(x_k)$ характеризует класс с меткой $+1$, количественно оцениваемая гистограммой с bin_j интервалами;

$D_{-1} = P[f(x_k) \in \text{bin}_j, y=-1]$ – степень достоверности того, что приведённое к диапазону $[0..1]$ значение составляющей признака $f(x_k)$ характеризует класс с меткой -1 , количественно оцениваемая гистограммой с bin_j интервалами.

Таким образом, на каждой итерации обучения алгоритма адаптивного усиления *AdaBoost* формируется слабый классификатор, обобщающий некоторую особенность изображений обучающей выборки.

Применение алгоритма адаптивного усиления *AdaBoost* к задаче автоматического обнаружения лиц на цифровых изображениях

Одним из успешных приложений метода адаптивного усиления *AdaBoost* является построение детекторов лиц на цифровых изображениях. Изображение лица фронтальной проекции при низком разрешении обладает хорошими обобщающими свойствами: на изображении лица имеются гладкие области в районе лба, щёк и носа и перепады яркости в районе глаз, бровей, основания носа и рта.

Механизм выделения обобщающих признаков изображения лица был предложен Viola и Jones [4]. Он базируется на использовании модифицированных фильтров Хаара. Важным свойством этих фильтров является простота вычислений, что определяет высокое быстродействие алгоритма при обучении слабых классификаторов. В алгоритме используются фильтры трёх типов с двумя прямоугольными областями, с тремя прямоугольными областями и с четырьмя прямоугольными областями. Их внешний вид представлен на рис. 1.



Рис. 1. Фильтры для выделения признаков лица

Значение отклика фильтра является результатом операции свёртки и представляет собой разность между суммами яркостей пикселей, заключенных в чёрных и белых прямоугольных областях соответственно. Несмотря на то, что фильтр может иметь различные геометрические размеры областей, вычисление значения отклика для фильтров одного типа происходит за одинаковое время. Это обеспечивается благодаря использованию интегрального изображения. Интегральное изображение в точке (X, Y) содержит сумму пикселей, расположенных слева и сверху этой точки и определяется как

$$I(X, Y) = \sum_{x' \leq X, y' \leq Y} I(x', y'),$$

где $I(X, Y)$ – входное изображение.

Интегральное изображение вычисляется за один проход по входному изображению:

$$S(X, Y) = S(X, Y-1) + I(X, Y)$$

$$I(X, Y) = I(X, Y-1) + S(X, Y),$$

где $S(X, Y)$ – кумулятивная строка суммы, $S(X, -1) = 0$, $I(-1, Y) = 0$.

В результате последовательной фильтрации входного изображения $I(X,Y)$ фильтрами различных типов, геометрических размеров и местоположения формируется набор признаков $x_i \in \{x_1, \dots, x_k, \dots, x_n\}$. В процессе обучения на обучающей выборке на каждой итерации слабый классификатор выделяет признак x_k , который способствует классификации образцов с наименьшей ошибкой на данной итерации.

Ниже представлены практические результаты обучения алгоритма *AdaBoost* для задачи автоматического обнаружения лиц на цифровых изображениях. В качестве обучающей выборки использовалась база данных из 2400 фронтальных изображений лиц и 2400 изображений, на которых лица отсутствуют, размерами 24x24 пикселей. Для каждого изображения путём обработки его набором фильтров сформирован вектор из 5445 признаков. На обучающей выборке в данном пространстве признаков обучались слабые классификаторы, построенные на базе классификатора байесовского типа со статистической оценкой плотности вероятности.

Процесс обучения алгоритма показан на рис. 2 и рис. 3.

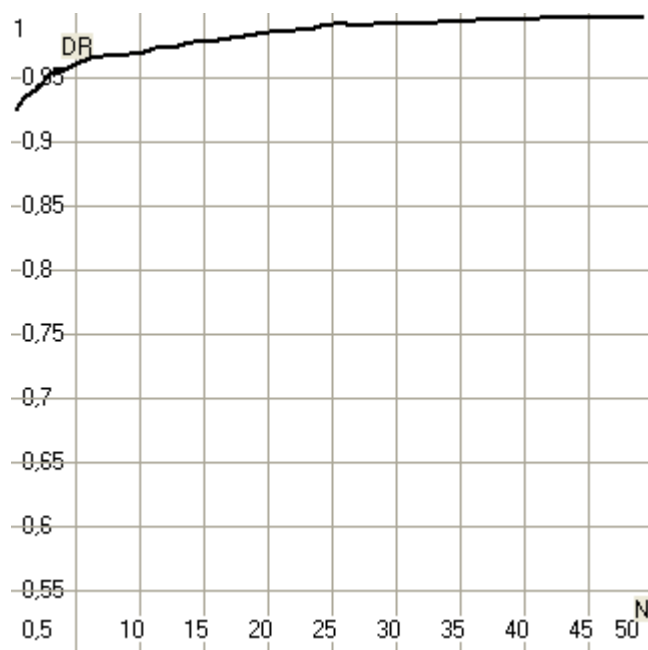


Рис. 2. Вероятность правильной классификации финальным классификатором в процессе обучения на обучающей выборке.

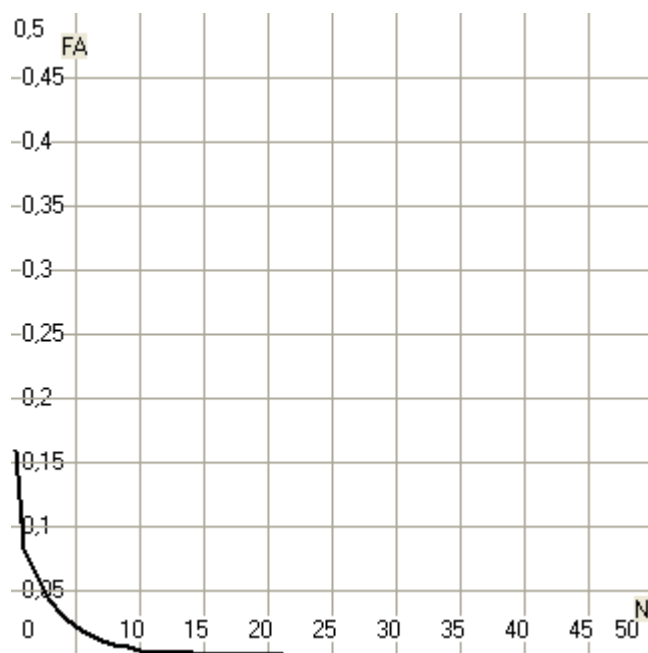


Рис. 3. Вероятность ложной классификации финальным классификатором в процессе обучения на обучающей выборке.

Как видно из рис. 2 обучение данного классификатора на обучающей выборке с использованием алгоритма *AdaBoost* позволяет достичь на 50-й итерации вероятности правильной классификации не менее 0,998.

График на рис. 3 наглядно демонстрирует такое свойство алгоритма *AdaBoost* как экспоненциальное уменьшение ошибки обучения финального классификатора по мере увеличения количества слабых классификаторов (числа итераций).

На рис.4 показаны фильтры признаков лица, которые по версии данного алгоритма обучения обладают наилучшими обобщающими свойствами.

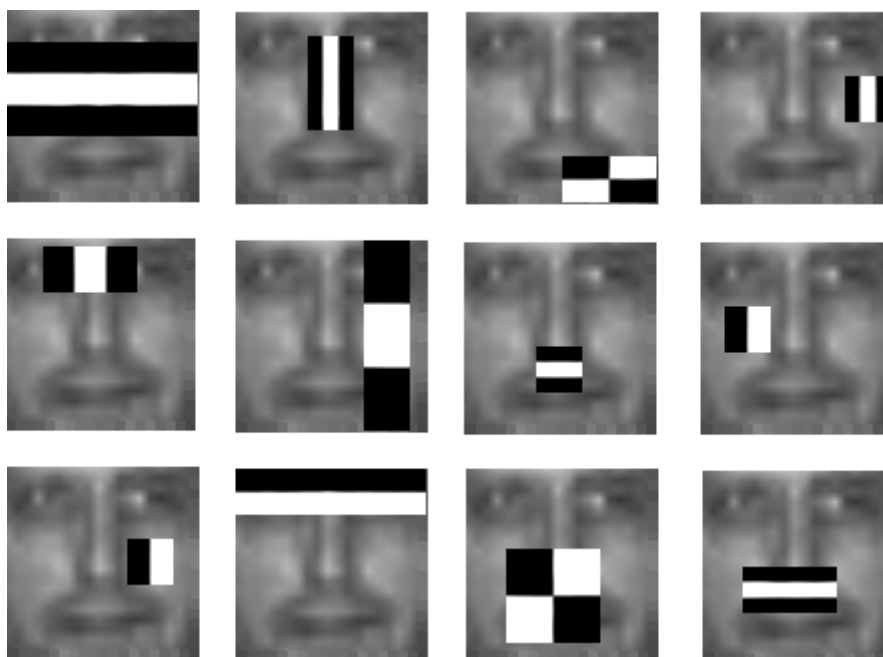


Рис. 4. Фильтры признаков лица, обладающих наилучшими обобщающими свойствами.

Работа алгоритма в режиме обнаружения осуществляется путём сканирования скользящим окном переменного масштаба всего поля изображения. Фрагмент изображения, выделенный скользящим окном, распознается (классифицируется) как лицо, либо как фоновое изображение. Примеры обнаружения изображений лиц с помощью разработанного алгоритма показаны на рис. 5 и 6.

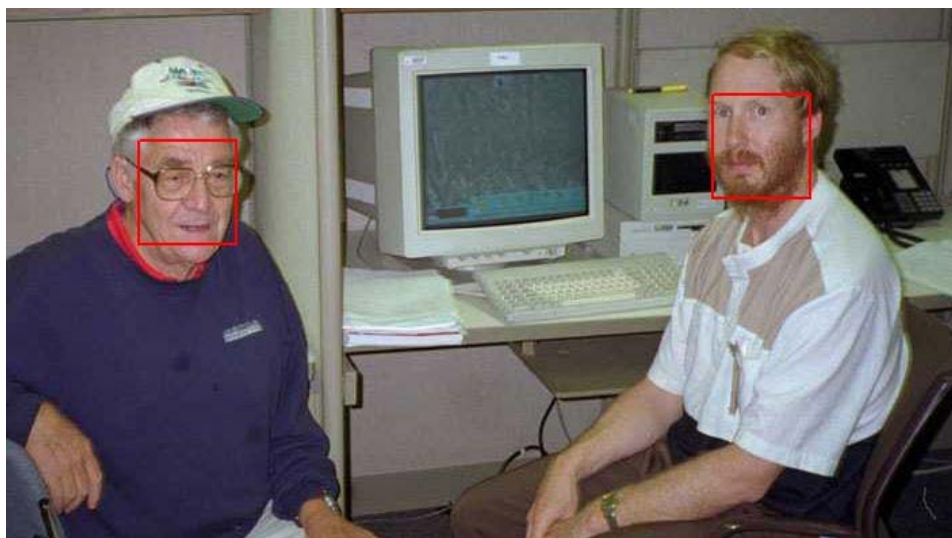


Рис. 5. Пример работы алгоритма (два лица в кадре)

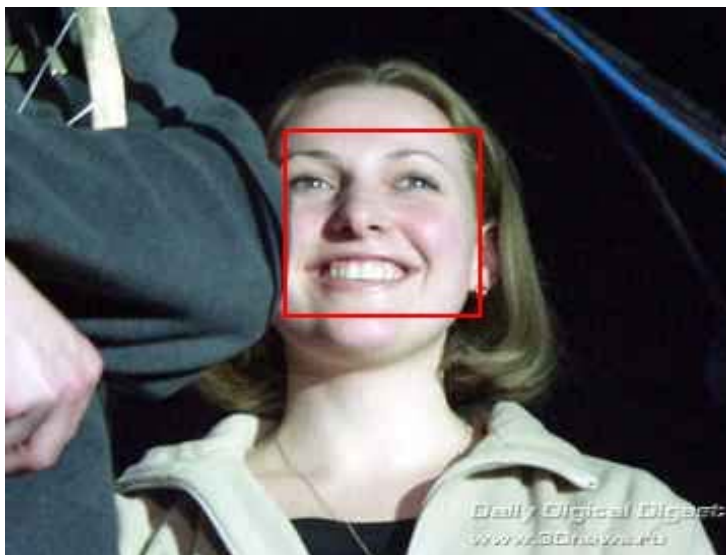


Рис. 6. Пример работы алгоритма (одно лицо в кадре)

Результаты проведённого математического моделирования показывают, что при работе по случайно выбранной совокупности тестовых изображений разработанный алгоритм автоматического обнаружения и локализации изображений лиц имеет следующие характеристики: вероятность правильного обнаружения и локализации не менее 0.95 при вероятности ложного обнаружения лица не более 0.01.

Заключение

В статье описаны результаты разработки алгоритмических и программных средств для автоматического обнаружения и локализации изображений лиц на цифровых изображениях. При построении алгоритмов использовался модифицированный алгоритм синтеза высокоэффективных классификаторов из множества слабых классификаторов – *Adaboost*, отличающийся от классического варианта использованием аппроксимации распределения вероятностей откликов слабых классификаторов вместо пороговой решающей функции. Данный подход позволил упростить алгоритм обучения классификатора, что привело к существенному сокращению времени обучения и повышению эффективности получаемого в результате сильного классификатора.

Проведенное математическое моделирование показало, что при работе по случайно выбранной совокупности тестовых изображений разработанный алгоритм автоматического обнаружения и локализации изображений лиц обеспечивает вероятность правильного обнаружения и локализации не менее 0.95 при вероятности ложного обнаружения лица не более 0.01. Кроме того, разработанный алгоритм обладает высокой производительностью, позволяющей работать с цифровым видеопотоком в реальном времени.

Разработанные алгоритмы могут быть использованы в системах биометрической идентификации, контроля доступа, в системах анализа и распознавания изображений, программных средствах для видеоконференций и других подобных приложениях.

Список использованных источников

1. Freund, Y., Schapire, R. A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55: 119-139, 1997
2. Freund, Y., Schapire, R. A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5): 771-780, September, 1999
3. Schapire, R., Freund, Y., Bartlett, P., Lee, W.S. Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. *The Annals of Statistics*, 26: 1651-1686, 1998
4. Viola, P., Jones, M. Robust Real Time Object Detection. *IEEE ICCV Workshop Statistical and Computational Theories of Vision*, July 2001