

УДК 004.931

ВЫДЕЛЕНИЕ ЛИЦА НА СНИМКЕ ИЗ ВИДЕОПОТОКА С ЦЕЛЬЮ ЕГО РАСПОЗНАВАНИЯ

Умяров Н.Х., Федяев О.И.

Донецкий национальный технический университет, Украина

В данной работе практически реализован один из важных этапов процесса автоматического распознавания лиц – локализация лица на снимках из видеопотока. Выполнен анализ современных методов поиска лица на изображении. Разработана программа выделения лиц на основе алгоритма Виола-Джонса.

1 Постановка задачи

В настоящее время задача распознавания человеческих лиц становится всё более актуальной. Процессу непосредственного распознавания лиц предшествует важный этап автоматической локализации лица на снимке, методы реализации которого сейчас активно развиваются. Поэтому целью данной работы является разработка такой подсистемы, которая вводит снимки с вэб-камеры, выделяет и локализует на них лица, сохраняет изображения лиц в базе данных для последующего распознавания [1].

Вводимый снимок $S(x,y)$ – представляет собой совокупность изображений отдельных объектов (лиц) и фона

$$S(x,y) = L_1(x,y) + L_2(x,y) + \dots + L_n(x,y) + F(x,y),$$

где n – количество выделяемых на снимке лиц; $F(x,y)$ – изображение фона; $(x,y) \in G$ – область определения снимка.

Снимок $S(x,y)$ с видеокамеры определён только на дискретном множестве точек D с координатами $(x_i, y_j) \in G$. При этом,

$$L_k(x_i, y_j) = \begin{cases} k, & \text{если } (x_i, y_j) \in D_k, \\ 0, & \text{иначе} \end{cases} \quad F(x_i, y_j) = 0, \text{ если } (x_i, y_j) \notin D_F$$

где D_k – дискретная область k -го лица; $D = D_1 \cup D_2 \cup \dots \cup D_n \cup D_F$; D – дискретная прямоугольная решётка, в узлах которой определён снимок $S(x,y)$. Узлы решётки определяются координатами (x_i, y_j) или их номерами (i,j) , в которых задаётся цвет снимка в данной точке, что в совокупности определяет пиксель снимка.

Таким образом, исходный снимок $S(x,y)$ можно рассматривать как прямоугольную целочисленную решётку P , узлы которой являются пикселями:

$$P = \{p(i, j) | i = \overline{1, N}, j = \overline{1, M}\},$$

где i, j – целочисленные координаты (номера) пикселя; $p(i, j) \in \{0, 1\}$ – цветовой компонент пикселя.

Задача работы заключалась в нахождении множества точек $(i,j) \in D_k$, для которых $S(x_i, y_j) = L_k(x_i, y_j)$.

2 Детектор лица по цвету кожи

Детектор цвета кожи, используя представление объектов в цвете, позволяет найти изображение человеческого лица на фотографии или в видеопотоке. В случае использования цветовой модели HSB для качественного детектирования необходимо правильно подобрать параметры цвета кожи.

Далее приведен алгоритм детектирования лица по цвету кожи:

```

 $\theta := 1;$  // Изначально задаётся высокий порог обнаружения лица
 $\Delta := 0.02;$  // Шаг уменьшения порога, если не найдено ни одно лицо
 $K := 0;$ 
 $S := 12;$  // Число сканируемых прямоугольников
ЦИКЛ ПОКА  $\theta > 0.5$ 

```

```

// Пройти всеми сканирующими прямоугольниками по изображению

$$\Pi_l^{pq} := \sum_{i=x_0}^{x_0+w_l} \sum_{j=y_0}^{y_0+h_l} P_{ij}, \quad l := \overline{1, C}, \quad p := \frac{M}{h_l}, \quad q := \frac{N}{w_l}$$

ЕСЛИ  $\frac{\Pi_l^{pq}}{w_l \times h_l} > \theta, \quad l := \overline{1, C}, \quad p := \frac{M}{h_l}, \quad q := \frac{N}{w_l}$ , ТО // Лицо найдено
K=K+1
// Отметить лицо на изображении
P = П(P, p, q, wl, hl),
где П(P, p, q, w, h) =  $\begin{cases} P_{ij} := RGB(255,0,0), \forall p_{ij}, q_{ij} : p_{11} < p_{ij} < w \ \& \ q_{11} < q_{ij} < h \\ P_{ij} := P_{ij}, \text{ иначе} \end{cases}$ ,
где  $i := \overline{1, w}, j := \overline{1, h}$ ;
RGB(r, g, b) – формирование пикселя на изображении
ЕСЛИ K < 1 ТО  $\theta := \theta - \Delta$ ;
ПОВТОРИТЬ;
```

Успешную работу алгоритма осложняет влияние освещённости на цвет кожи выделяемого лица. На рис. 1 показаны два изображения: на верхнем изображении лицо выделилось успешно, а на нижнем – неудачно. Следовательно, недостаток данного алгоритма – сильная зависимость от освещения.

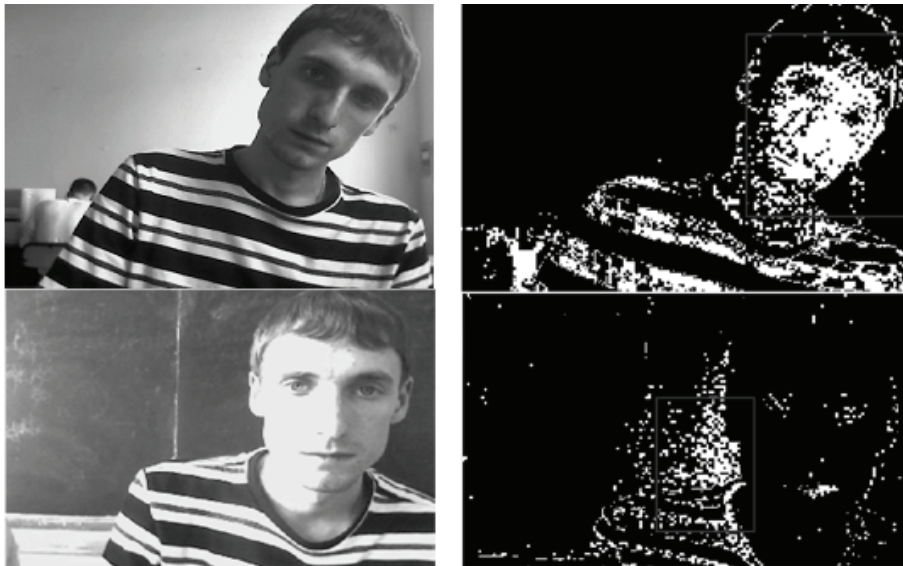


Рисунок 1. Результаты выделения лица при разной освещённости изображения

3 Алгоритм Виола-Джонса

Как видно, алгоритм выделения лица по цвету кожи не годится для формирования качественной обучающей выборки. Поэтому были рассмотрены другие методы, которые опирались на антропологические особенности лица – расположение глаз, носа, рта, ушей. Из этого класса наилучшие рекомендации имел алгоритм Виола-Джонса, описанный в 2001 году [2].

В его основе лежит алгоритм adaptive boosting'a (адаптивного усиления) или сокращённо AdaBoost. Смысл алгоритма заключается в том, что если есть набор эталонных объектов, т.е. есть изображение и класс к которому оно принадлежит (например, -1 – нет лица на изображении, +1 – есть лицо). Кроме того, имеется множество простых классификаторов, из которых можно построить один более совершенный и мощный классификатор. При этом в многошаговом процессе составления и обучения финального классификатора акцент делается на эталоны, которые распознаются «хуже». В этом и заключается адаптивность алгоритма, т.к. в процессе обучения он подстраивается под наиболее «сложные» объекты. Примеры его применения показали, что AdaBoost очень эффективный и быстрый алгоритм. Он хорошо себя зарекомендовал при решении таких задач, в которых выделяемые объекты (например, лица) сильно зашумлены. На рис. 2 показана идея локализации лица на изображении по данному алгоритму.



Рисунок 2. Идея алгоритма Виола-Джонса из библиотеки OpenCV

В основе алгоритма лежат три ключевые концепции:

а) Использование «Интегральных изображений», позволяющие вычислять площадь окна сканирования в несколько раз быстрее, чем в алгоритме детектирования по цвету кожи (см. подраздел 2). Интегральное изображение в точке (x, y) включает в себя сумму пикселей, которые расположены левее и выше изображения:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \quad (1)$$

где $ii(x, y)$ – значение интегрального изображения в точке (x, y) .

Для вычисления интегрального изображения используются следующие рекуррентные формулы:

$$\begin{aligned} s(x, y) &= s(x, y-1) + i(x, y) \\ ii(x, y) &= ii(x-1, y) + s(x, y) \end{aligned} \quad (2)$$

где $s(x, y)$ – накапливаемая сумма строки, $s(x, -1) = 0$, $ii(-1, y) = 0$.

б) Адаптивный алгоритм обучения выбирает небольшое число признаков низкого уровня из более чем 180 тыс. признаков. Существует много мотиваций для использования признаков, а не отдельных пикселей. Самая важная причина: признак может кодировать разные значения, которые трудно определять и изучать, используя конечное множество обучаемых данных. Признаки вычисляются намного быстрее, чем пиксельные величины, что подробно будет рассмотрено далее. В алгоритме Виола-Джонса используется три вида признаков: два прямоугольника, три прямоугольника и четыре, хотя в действительности число видов признаков может быть и больше. Область глаз более тёмная по сравнению со средней областью лица и переносицы. Прimitives этой конфигурации и размеров наиболее лучшим образом «характеризуют» данное изображение. На основе таких классификаторов с отобранными наиболее эффективными примитивами строится каскад классификаторов. Каждый последующий элемент каскада имеет более жёсткие условия успешного прохождения, чем предыдущий (используется больше примитивов). Тем самым до конца доходят только самые «правильные». Обучение базируется на идее бустинга. Для успешного обучения требуется несколько тысяч «положительных» и «отрицательных» изображений (в зависимости от того, есть ли на нём лицо, или нет). Далее показан алгоритм бустинга [3]:

1) Дана обучающая выборка в формате $(x_1, y_1), \dots, (x_n, y_n)$, где

$$w_{l,i} = \begin{cases} 0.5/m, & \text{если } y_i = 0 \\ 0.5/l, & \text{если } y_i = 1 \end{cases}$$

Инициализировать веса: $y_i = \begin{cases} 1, & \text{face is in the picture} \\ 0, & \text{other} \end{cases}$

где m и l – число «положительных» (есть лицо на снимке) и «отрицательных» изображений (нет лица на

снимке).

Для $t = 1, \dots, T$

2) Нормализовать веса: w_t – вероятностно распределены.

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

Для каждого признака j тренировать классификатор h_j , который ограничен использованием одного признака.

Ошибка обучения ε_j при этом находится относительно w_t :

$$\varepsilon_j = \sum_i w_i |h_j(x_i) - y_i|$$

а) Выбрать классификатор h_t с наименьшей ошибкой ε_j .

б) Обновить веса:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-\varepsilon_i},$$

где $\varepsilon_j = 0$, если пример x_i классифицировался корректно; $\varepsilon_j = 1$, иначе и $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.

с) Результирующий классификатор будет равен:

$$h(x) = \begin{cases} 1, & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t, \text{ где } \alpha_t = \log \frac{1}{\beta_t} \\ 0, & \text{ иначе} \end{cases}$$

Бустинг используется для выбора небольшого конечного множества признаков высокого уровня и для тренировки классификатора. Для обучения нейросетевого классификатора применяется простой (слабый, англ. weak) алгоритм обучения.

в) Использование каскада Хаара [4], который выбирает небольшое число признаков и объединяет их в признаки более высокого уровня. В результате получаем каскад из 38 шагов на большом наборе данных, состоящем из 507 лиц с более чем 75 миллионами сканирующих окон. Лица распознаются со скоростью 10 вычислений признаков в секунду для каждого окна. Результирующий детектор скользит по изображению в разных масштабах, при этом увеличивается не масштаб изображения, а масштаб самого окна. Таким образом приобретается инвариантность к отдалению лица от камеры.

4 Программная реализация

Программный продукт автоматического выделения лица на снимках видеопотока был реализован на ноутбуке со встроенной веб-камерой и операционной системой семейства Windows. Так как информация, передаваемая с веб-камеры, поступает в виде потока байтов, то для их обработки использовался стандартный Microsoft WebCam driver, а также библиотеки – надстройки, реализующие работу с веб-камерой на более высоком уровне. Всё это существенно облегчило решение поставленной задачи, благодаря избавлению от прямого вызова низкоуровневых функций. Результат работы программы выделения лиц показан на рис. 3.

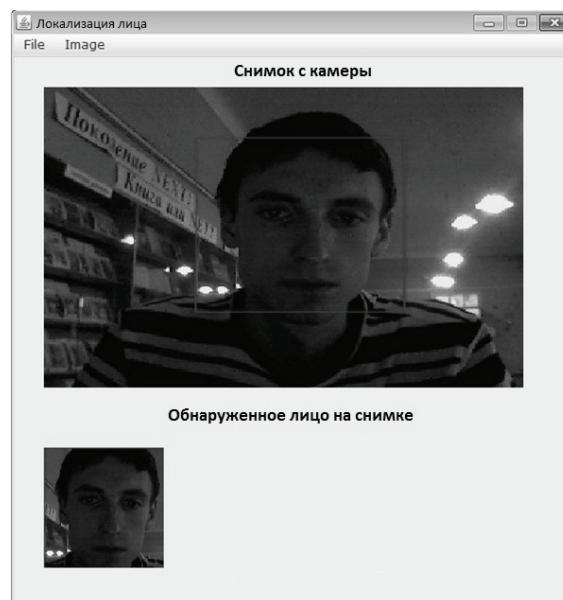


Рисунок 3. Результат автоматического выделения лица методом Виола-Джонса

Выводы

В работе проанализированы современные алгоритмы выделения лиц на снимках, принимаемых из веб-камеры в реальном времени. Изучены основные особенности цветового формата HSB. Построен алгоритм распознавания лица по цвету кожи человека. Из-за чувствительности этого алгоритма к изменению яркости изображения предпочтение было отдано методу Виола-Джонса. Выполнена программная реализация алгоритма Виолы-Джонса, позволившая эффективно выделять лица на компьютерных изображениях.

Выполненная работа является важным этапом на пути построения системы распознавания лица человека в режиме реального времени.

Литература

- [1] Умяров Н.Х., Федяев О.И. Параметрическая модель свёрточной нейронной сети // VI международная научно-техническая конференция студентов, аспирантов и молодых научных работников «Информатика и компьютерные технологии»: Т. 2 – Донецк, ДонНТУ, 2010. – 292с.
- [2] Paul Viola, Michael Jones. Rapid Object Detection using a Boosted Cascade of Simple Features, 2001. - 8 pages
- [3] Саймон Хайкин. Нейронные сети: полный курс.: издательский дом «Вильямс», 2006. - 1104с.
- [4] Rainer Lienhart, Alexander Kuranov, Vadim Pisarevsky. Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection.: Intel