

УДК 004.93'1

Г. Б. Южаков

ФГУП ЦНИИХМ

Московский физико-технический институт (государственный университет)

Алгоритм быстрого построения дескрипторов изображения, основанных на технике гистограмм ориентированных градиентов

Рассматривается задача составления дескрипторов прямоугольных фрагментов изображения в задаче поиска объектов на изображениях. При этом предлагается использовать технику, основанную на гистограммах ориентированных градиентов. В работе описываются методы оптимизации скорости расчетов, такие как техника построения интегральных изображений, прореживание, метод имитации скольжения и техника вспомогательных дескрипторов. В заключение приводится алгоритм реализации основных типов основанного на гистограммах ориентированных градиентов описания изображений. Вычислительный эксперимент показал эффективность описанных в работе методов.

Ключевые слова: гистограмма ориентированных градиентов, интегральное изображение, интегральные бины, скользящее окно, имитация скольжения, прореживание, вспомогательные дескрипторы.

1. Введение

Описание изображения является промежуточным, но крайне важным этапом решения задач компьютерного зрения. В процессе описания изображения производится преобразование визуальных данных, содержащихся на изображении, в приемлемую для алгоритма классификации форму.

Так как в визуальных данных обычно содержится много излишней информации, то задача описания помимо прочего состоит в исключении как можно большего количества неинформативных признаков при сохранении данных обо всех существенных для задачи распознавания характеристиках исследуемого фрагмента изображения.

В данной статье исследуется описание изображений в задаче распознавания, и представлена реализация быстрого алгоритма построения класса дескрипторов, основанных на технике гистограмм ориентированных градиентов (HOG – histogram of oriented gradients).

HOG дескрипторы используются классификаторами SVM, Random Forest, Boosting и многими другими.

2. Гистограмма ориентированных градиентов

В основе данного метода лежит предположение, что вид распределения градиентов интенсивности изображения позволяет достаточно точно определить наличие и форму присутствующих на нем объектов.

При описании фрагмента изображения оно разбивается на несколько небольших участков, далее называемых *ячейками*. В ячейках вычисляются гистограммы h_i направленных градиентов внутренних точек. Обычно они объединяются в одну гистограмму $h = f(h_1, \dots, h_k)$, после чего она нормализуется по яркости (L_2 или L_1 норма):

$$h_{L_2} = \frac{h}{\sqrt{|h|_2^2 + \varepsilon}}, h_{L_1} = \frac{h}{|h|_1 + \varepsilon}, h_{\sqrt{L_1}} = \sqrt{h_{L_1}},$$

где ε — некоторая малая константа.

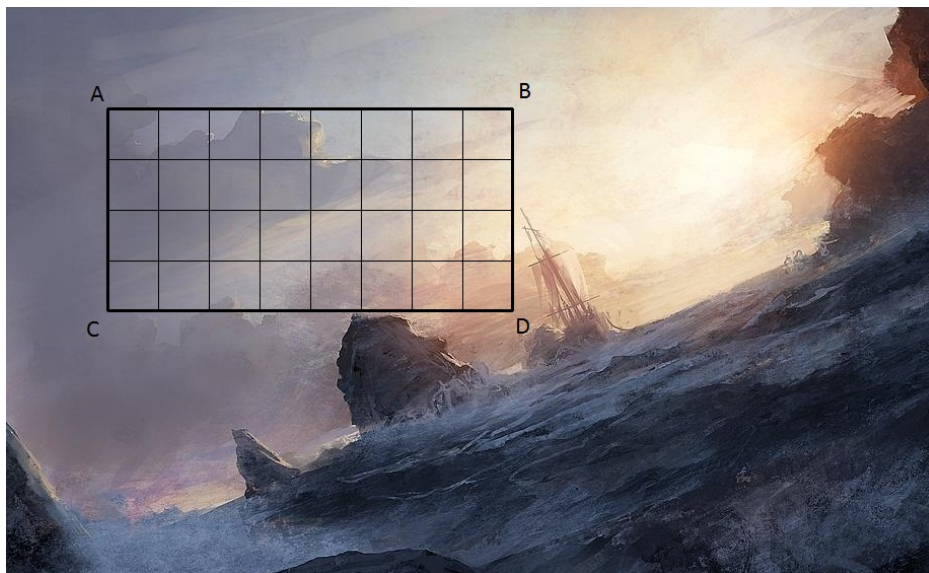


Рис. 1. Скользящее окно

Таким образом, данный описатель содержит пространственную информацию о фрагменте и инвариантен к освещению.

При вычислении градиентов производится свертка изображения с ядрами $[-1, 0, 1]$ и $[-1, 0, 1]^T$, в результате чего образуются две матрицы D_x и D_y производных вдоль осей x и y соответственно. Эти матрицы используются для вычисления углов и величин (модулей) градиентов в каждой точке изображения.

Пусть множество углов $(-\pi, \pi]$ разбивается на n равных интервалов вида $(-\frac{k-1}{n}\pi, \frac{k}{n}\pi]$, где $k = \{1, \dots, n\}$. Каждому интервалу ставится в соответствие бин гистограммы. Тогда гистограмма ячейки заполняется так, что величина градиента в каждой внутренней его точке добавляется к величине бина, соответствующего интервалу, содержащему угол данного градиента.

Обычно ячейки, так же, как и сами фрагменты, имеют прямоугольную форму (рис. 1), которая позволяет применять технику интегральных изображений.

3. Вспомогательные техники

Алгоритм построения дескрипторов удобно рассматривать как систему, состоящую из нескольких вспомогательных техник:

- 1) интегральное представление изображений;
- 2) интегральные градиенты;
- 3) прореживание;
- 4) имитация алгоритма скольжения;
- 5) вспомогательные дескрипторы.

Рассмотрим их более подробно.

3.1. Интегральное представление изображений

Во многих задачах обработки изображений, таких как построение гистограмм ориентированных градиентов и фильтры Хаара, необходимо вычислять суммарную яркость прямоугольных участков изображения. Техника интегральных изображений позволяет существенно ускорить данную операцию. При этом время вычисления суммы яркости прямоугольника не зависит от его площади.

Пусть имеется отображение I из множества матриц $\mathbb{R}^{m \times n}$ в $\mathbb{R}^{(m+1) \times (n+1)}$, такое что

$$\forall X \in \mathbb{R}^{m \times n} \exists Y \in \mathbb{R}^{(m+1) \times (n+1)}, Y = I(X) :$$

$$Y(x, y) = \begin{cases} \sum_{i=0, j=0}^{x-1, y-1} X(i, j), & \text{при } x > 0 \text{ и } y > 0; \\ 0, & \text{при } x = 0 \text{ или } y = 0. \end{cases}$$

Расчет матрицы $Y = I(X)$ занимает линейное время, пропорциональное числу пикселей в исходном изображении:

$$Y(x, y) = X(x-1, y-1) - Y(x-1, y-1) + Y(x, y-1) + Y(x-1, y).$$

Время ее вычисления можно немного уменьшить, если сначала проинтегрировать столбцы матрицы X , а потом строки. При этом время вычисления интегрального изображения размером 500×500 уменьшается с 1,242 мс до 1,108 мс.

По интегральной матрице (рис. 2) всего за 4 обращения можно вычислить сумму интенсивностей пикселей внутри произвольного прямоугольника $ABCD$ исходного изображения X . Суммарная интенсивность в прямоугольнике $ABCD$ вычисляется по формуле

$$S_X(ABCD) = Y(A_x, A_y) + Y(C_x, C_y) - Y(B_x, B_y) - Y(D_x, D_y).$$

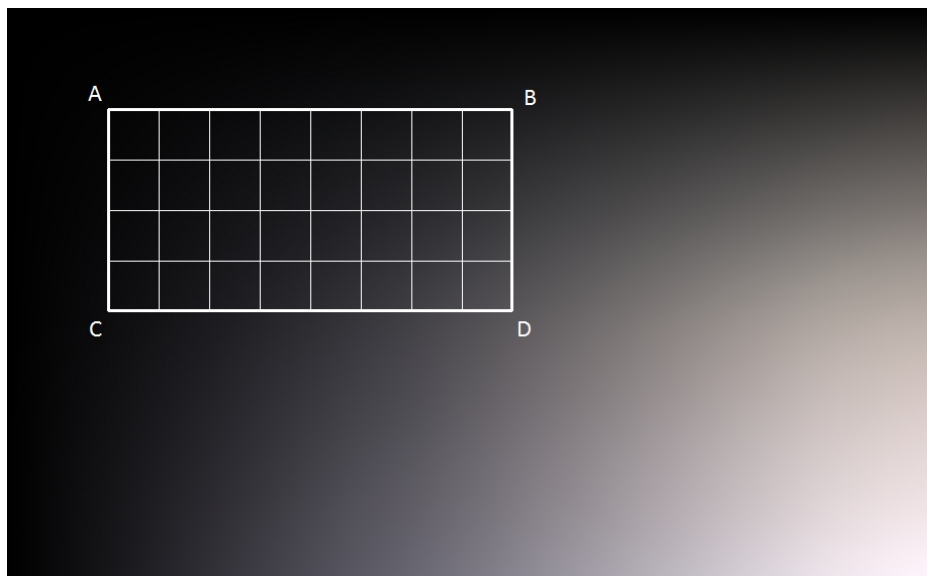


Рис. 2. Интегральное изображение

Более эффективно совершать подобное вычисление сразу для всего изображения. Пусть имеется матрица $X \in \mathbb{R}^{m \times n}$, тогда суммирование S_{wh} изображения X по множеству всех прямоугольников размером $w \times h$ задается следующим образом:

$$\forall X \in \mathbb{R}^{m \times n} \exists Z \in \mathbb{R}^{(m-w+1) \times (n-h+1)}, Z = S_{wh}(X) :$$

$$Z(x, y) = Y(x+w, y+h) + Y(x, y) - Y(x, y+h) - Y(x+w, y),$$

где $Y = I(X)$, $x \in \{0, \dots, m-w\}$, $y \in \{0, \dots, n-h\}$.

В статье техника интегральных изображений используется для построения интегральных градиентов изображения, с помощью которых вычисляются НОГ дескрипторы.

3.2. Интегральные градиенты

Пусть имеется две разностные матрицы D_x и D_y исходного изображения $X \in \mathbb{R}^{m \times n}$ вдоль осей x и y соответственно:

$$D_x(x, y) = X(x, y) - X(x + 1, y), x \in \{0, \dots, m - 1\}, y \in \{0, \dots, n\},$$

$$D_y(x, y) = X(x, y) - X(x, y + 1), x \in \{0, \dots, m\}, y \in \{0, \dots, n - 1\}.$$

С их помощью вычисляются матрицы углов A и модулей V градиентов:

$$A(x, y) = \arctan \frac{D_x(x, y)}{D_y(x, y)}, V(x, y) = \sqrt{D_x(x, y)^2 + D_y(x, y)^2},$$

где $x \in \{0, \dots, m - 1\}, y \in \{0, \dots, n - 1\}$.

Для построения гистограммы из n бинов эти матрицы приводятся к дискретному виду:

$$A_n(x, y) = [A(x, y) \frac{n}{2\pi}],$$

$$V_i(x, y) = \begin{cases} V(x, y), & \text{если } A_n(x, y) = i; \\ 0, & \text{если } A_n(x, y) \neq i, \end{cases}$$

где n — количество бинов, $i \in \{0, \dots, n - 1\}$, $[x]$ — целая часть x . Тогда бинами интегральных градиентов изображения $B(X)$ будет называться множество матриц $I(V_i)$.

Экспериментально было выявлено [1], что оптимальное качество распознавания достигается при количестве бинов порядка восьми.

3.3. Прореживание

Вычисление дескрипторов осуществляется в процессе сканирования изображения на разных масштабах некоторым скользящим окном U (рис. 1). Пусть $U = (U_x, U_y)$ — его текущие размеры, $dU = (dU_x, dU_y)$ — шаг скольжения, $V = (V_x, V_y)$ — размер ячеек, из которых оно состоит, а $dV = (dV_x, dV_y)$ — расстояние между ними.

Введем также понятие единицы длины $E = (E_x, E_y)$, как вектор наибольшей длины, которому кратны векторы U, dU, V и dV . При $E \neq (1, 1)$, не влияя на окончательный результат, интегральное изображение может быть прорежено так, чтобы новая единица длины стала равна единичному вектору. Прореженная матрица $T(Y)$ интегрального изображения $Y = I(X)$ вычисляется следующим образом:

$$\forall Y \in \mathbb{R}^{(m+1) \times (n+1)} \exists Z \in \mathbb{R}^{(1+[m/E_x]) \times (1+[n/E_y])}, Z = T(Y) :$$

$$Z(u, v) = Y(u \cdot E_y, v \cdot E_x).$$

При прореживании все параметры скольжения сохраняют свои значения при всех масштабах скользящего окна, меняется только единица длины E . При этом имеют постоянные значения отношения $\frac{dU_x}{U_x}$ и $\frac{dU_y}{U_y}$, поэтому качество распознавания инвариантно к размеру скользящего окна.

В статье прореживание применяется к бинам интегральных градиентов.

3.4. Имитация алгоритма скольжения

Однотипные операции сразу со всеми элементами изображения, имеющими представление в виде числовой матрицы, поддаются оптимизации, поэтому скорость их выполнения может в несколько раз превосходить скорость поэлементного вычисления. Именно на этом основана техника имитации алгоритма скольжения.

Пусть рабочая матрица X имеет размеры $m \times n$. Тогда у скользящего окна U , размером $w \times h$, имеется сопряженное скользящее окно U' (рис. 3), размером $(m - w + 1) \times (n - h + 1)$.

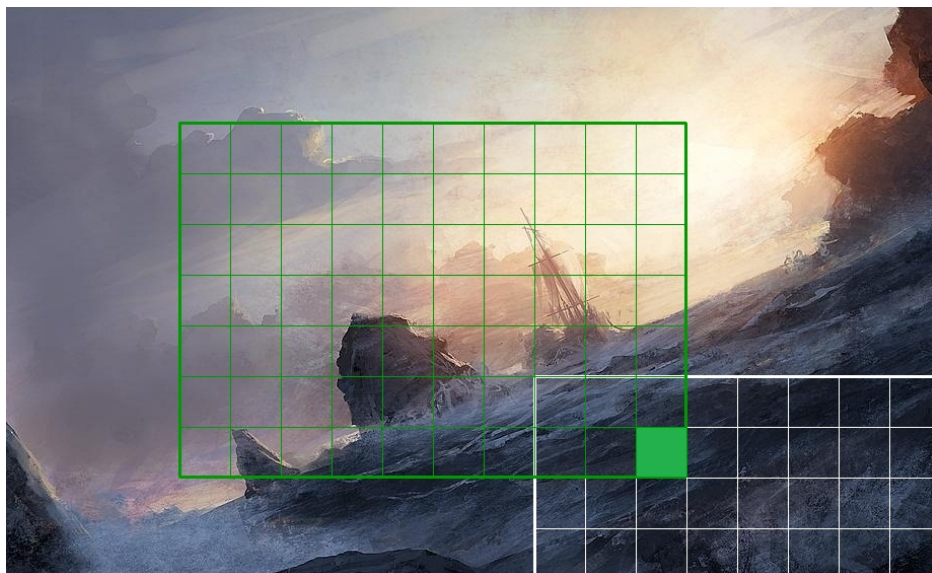


Рис. 3. Сопряженное скользящее окно (зеленый прямоугольник). Клетка в правом нижнем его углу соответствует извлекаемому элементу данных с координатами как в белом прямоугольнике

Таким образом, сканирование матрицы X скользящим окном U можно заменить сканированием окном U' .

Пусть шаг скольжения $dU = (1, 1)$, тогда сопряженное окно будет скользить с шагом $dU' = dV$. Положение окна U' может быть вычислено по формуле

$$U'_x(u, v) = u \cdot dU'_x, U'_y(u, v) = v \cdot dU'_y,$$

где $u \in \{0, \dots, [w/dU'_x]\}$, $v \in \{0, \dots, [h/dU'_y]\}$.

В каждом своем положении (u, v) окно U' покрывает некоторую подматрицу X_{uv} матрицы X :

$$X_{uv}(x, y) = X(x + U'_x(u, v), y + U'_y(u, v)),$$

где $x \in \{0, \dots, m - w + 1\}$, $y \in \{0, \dots, n - h + 1\}$.

Подматрица X_{uv} вытягивается в строчку и добавляется в матрицу искомым дескрипторов D .

Введем операцию R_{mn} перевода матрицы размера $m \times n$ в матрицу размером $(m+n) \times 1$, такую что

$$\forall X \in \mathbb{R}^{m \times n} \exists Y \in \mathbb{R}^{(m+n) \times 1},$$

$$Y = R_{mn}(X), Y(u, v) = X(x, y),$$

где $x = u - y \cdot m$, $y = [u/m]$, $u = \{0, \dots, (m+n) - 1\}$, $v = 0$. Тогда

$$D(\cdot, v \frac{w}{dU'_x} + u) = R_{mn}(X_{uv}).$$

Аналогично задается обратная операция R_{mn}^{-1} :

$$\forall Y \in \mathbb{R}^{(m+n) \times 1} \exists X \in \mathbb{R}^{m \times n},$$

$$X = R_{mn}^{-1}(Y), X(x, y) = Y(u, v),$$

где $u = x + y \cdot m$, $v = 0$, $x = \{0, \dots, m - 1\}$, $y = \{0, \dots, n - 1\}$.

Дескрипторы будут содержаться в столбцах матрицы D . При этом по номеру столбца можно однозначно определить размер и положение описываемого им фрагмента изображения.

Пусть теперь $dU \neq (1, 1)$, тогда матрица X делится на $dU_x dU_y$ одинаковые части X_{ij} , размером $(m/dU_x) \times (n/dU_y)$ каждая:

$$X_{ij}(x, y) = X(x \cdot dU_x + i, y \cdot dU_y + j),$$

где $i = \{0, \dots, dU_x - 1\}, j = \{0, \dots, dU_y - 1\}$. Тогда окно U' будет скользить по множеству матриц X_{ij} . Выбор матрицы из этого множества (i, j) зависит от положения окна (u, v) :

$$i = (u \cdot dV_x) \bmod dU_x, j = (v \cdot dV_y) \bmod dU_y.$$

Остальные параметры скольжения изменятся следующим образом:

$$dU'_x = [dV_x/dU_x], dU'_y = [dV_y/dU_y], w' = [w/dU_x],$$

$$h' = [h/dU_y], m' = [m/dU_x], n' = [n/dU_y].$$

3.5. Вспомогательные дескрипторы

В более общем случае вычисление дескрипторов производится в несколько этапов скольжения. Пусть полученные на промежуточном этапе скольжения дескрипторы D преобразуются в матрицу D' некоторым оператором F :

$$\forall D \in \mathbb{R}^{(m+n) \times h} \exists D' \in \mathbb{R}^{(m+n) \times h'}, D' = F(D).$$

Из строк матрицы D' составляются матрицы D'_i , по которым совершается независимое скольжение:

$$D'_i \in \mathbb{R}^{m \times n}, D'_i = R_{mn}^{-1}(D'(\cdot, i)), i = \{1, \dots, h'\}.$$

Когда все этапы скольжения пройдены, эти матрицы дескрипторов объединяются в одну итоговую матрицу.

4. Конкретные описатели

Рассмотрим особенности реализации нескольких основных типов описания изображений, основанных на технике гистограмм ориентированных градиентов.

4.1. Характеристика SHOG

SHOG (simple histogram of oriented gradients) описатель является наиболее простым из NOG описателей. Если X – исходное изображение, то алгоритм его построения записывается следующим образом:

- 1) вычисляются интегральные бины изображения $\{B_i\} = B(X), i = \{1, \dots, n\}$, где n – количество бинов;
- 2) если единица длины $E \neq (1, 1)$, то бины прореживаются $B_i = T(B_i)$;
- 3) бины B_i суммируются по размеру ячейки V скользящего окна, $X_i = S_{V_x V_y}(B_i)$;
- 4) в результате скольжения по X_i , вычисляются матрицы дескрипторов D_i .
- 5) вычисленные дескрипторы D_i объединяются в искомую матрицу дескрипторов D .

Приведенные ниже виды описания включают этот алгоритм как составную часть.

4.2. Характеристика РНОГ

Описатель РНОГ (pyramid histogram of oriented gradients) [2] отличается от простого тем, что состоит из нескольких независимых этапов скольжения с разным размером ячеек V скользящего окна. Каждому размеру его ячеек соответствует так называемый уровень пирамиды. Обычно размеры ячеек на разных уровнях соотносятся как степень двойки:

$$V_x^i = U_x/2^i, V_y^i = U_y/2^i, i = \{0, \dots, L\}, L \leq \lceil \log \min(U_x, U_y) \rceil.$$

Рассмотрим отдельно k -й уровень пирамиды. Вклад, который он внесет в квадрат модуля вектора дескриптора по отношению к нулевому уровню, составит величину порядка $\sum_{i=1}^{4^k} (4^{-k})^2 = 4^{-k}$. Чтобы вклад каждого уровня был одного порядка, необходимо умножить все элементы k -го уровня на 2^k .

В силу резкого роста вычислительных затрат и падения важности элементов дескриптора с увеличением номера уровня пирамиды в большинстве прикладных задач имеет смысл использовать уровни не выше третьего.

4.3. Характеристика ОНОГ

Описатель ОНОГ (original histogram of oriented gradients) был предложен в одной из первых статей [1], в которой изложена техника гистограмм ориентированных градиентов. В нем учитывается локальная зависимость нескольких соседних ячеек друг от друга. Иначе говоря, используется техника вспомогательных дескрипторов, которые нормируются по L_2 норме.

4.4. Характеристика ГНОГ

Можно объединить преимущества нескольких дескрипторов, например, ОНОГ и РНОГ, создав из них композицию. В этом и состоит суть ГНОГ (general histogram of oriented gradients) описателя. Пусть d – ГНОГ дескриптор, d_1 – РНОГ дескриптор длины m , d_2 – ОНОГ дескриптор длины n . Тогда

$$d = \frac{1}{2}(d_{11}, \dots, d_{1m}, d_{21}, \dots, d_{2m}), \sum_{i=1}^m d_{1i} = 1, \sum_{i=1}^n d_{2i} = 1, \sum_{i=1}^{m+n} d_i = 1.$$

Аналогичным образом создаются композиции произвольного числа k нормированных на единицу дескрипторов:

$$d_{i1} = (d_{i1}, \dots, d_{in_i}), \sum_{j=1}^{n_i} d_{ij} = 1, d = \frac{1}{m}(d_{11}, \dots, d_{1n_1}, \dots, d_{k1}, \dots, d_{kn_k}).$$

5. Экспериментальные результаты

В вычислительном эксперименте тестовое изображение имело размер 1000×500 , начальное значение единицы длины было равно 10×10 , пропорции скользящего окна 4×4 , шаг скольжения 1×1 (в единицах длины). В таблице 1 приведено потраченное на составление дескрипторов время. Результаты получены на процессоре Intel Core i7, 2.8 GHz. В случае многомасштабного скольжения с коэффициентом роста масштаба 1,2 время скольжения увеличивается примерно в 3 раза.

Время построения дескрипторов всего изображения

Тип описателя	Параметры скольжения	Время скольжения (мс)	Время многомасштабного скольжения (мс)
SHOG	$L = 3$	23,7	66,2
RHOG	$L = \{1, 2, 3\}$	44,2	119,0
OHOG	$C = 2 \times 2, dC = 1 \times 1$	11,3	27,5
GHOG	$L = \{1, 2, 3\},$ $V = 2 \times 2, dV = 1 \times 1$	55,1	147,3

6. Вывод

Таким образом, в работе был приведен алгоритм быстрого описания всего изображения, основанного на технике гистограмм ориентированных градиентов, инвариантного к геометрическим и фотометрическим искажениям, к степени освещенности и к размеру скользящего окна. При этом время построения дескрипторов — порядка 100 мс. Этой скорости вполне достаточно, чтобы использовать алгоритм как часть многих прикладных задач в распознавании образов и компьютерном зрении.

Литература

1. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection // INRIA. — 2005.
2. Bosch A., Zisserman A., Munoz X. Representing shape with a spatial pyramid kernel // CIVR. — 2007.

Поступила в редакцию 10.06.2012.