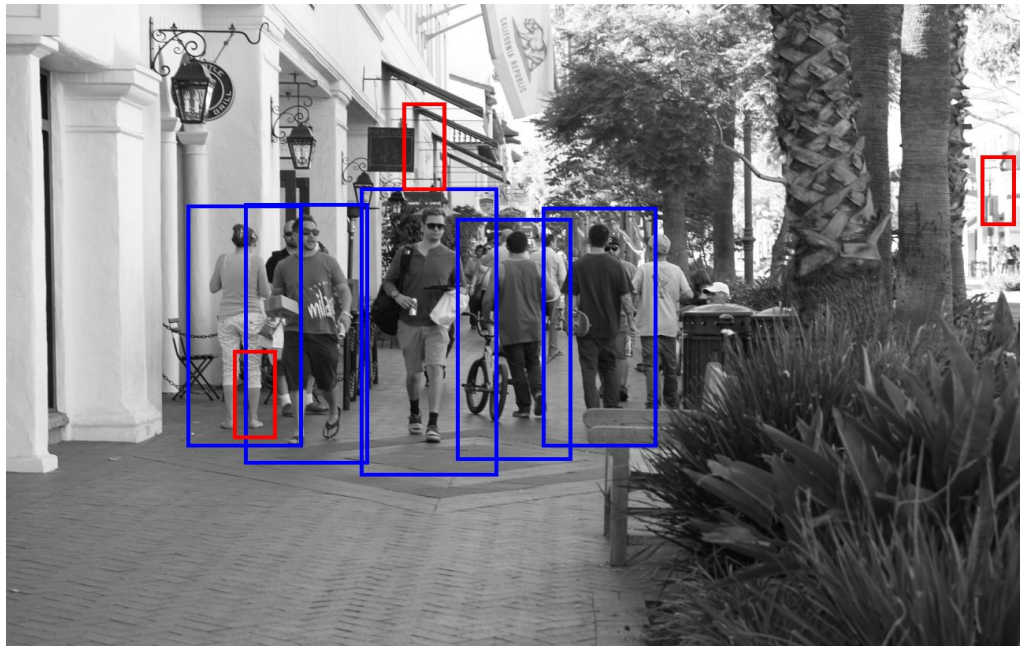


Гистограммы ориентированных градиентов(НОГ) для распознавания людей на изображениях.



Стадии получения HOG дескриптора из изображения:

- 1) Нормализация гаммы и цвета
- 2) Вычисление градиентов
- 3) Взвешенное распределение градиентов в гистограмму ориентаций
- 4) Контрастная нормализация пересекающихся пространственных блоков
- 5) Построение HOG дескриптора по всему распознающему окну.
- 6) Линейная классификация человек/не человек с помощью SVM.

Идея работы алгоритма

Основная идея этого алгоритма в том, что локальная структура изображения может хорошо описываться распределением локальных градиентов интенсивности, даже без точной информации о их местоположении.

На практике это реализовано разбиением изображения на маленькие клетки. Именно в этих клетках будут вычисляться градиенты для каждого пикселя, которые потом будут распределены по соответствующим столбцам гистограммы ориентаций.

Также необходима хорошая локальная нормализация.

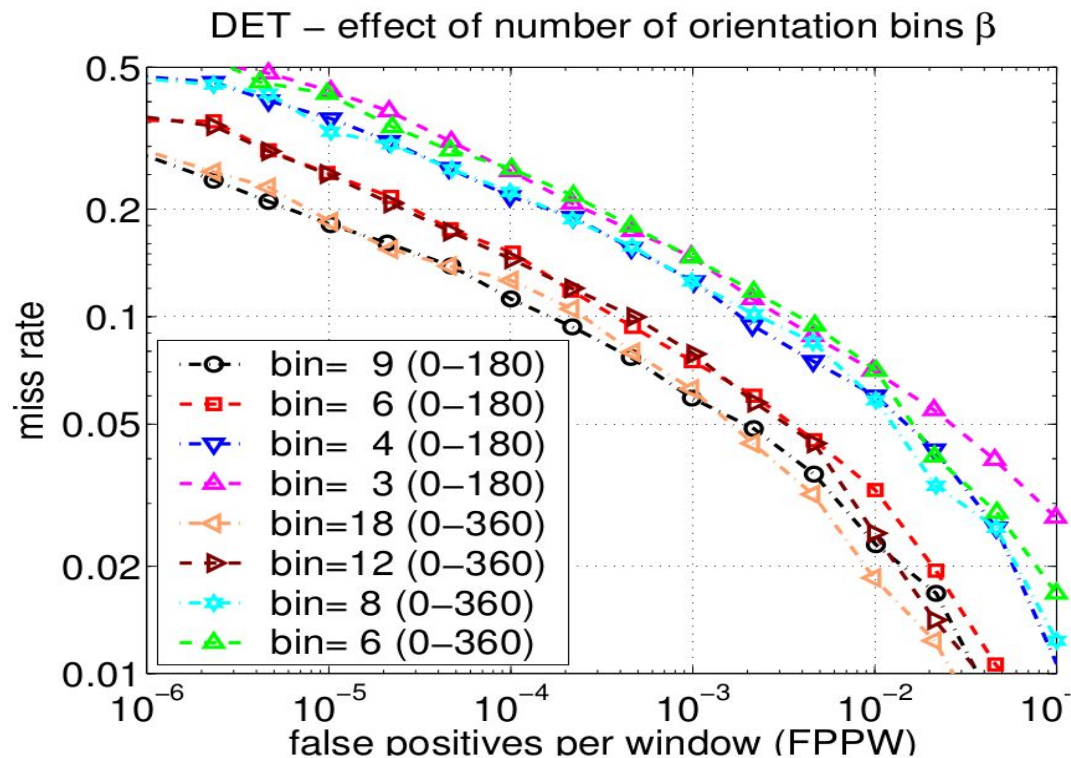
Образцы из набора изображений



Особенности проведения эксперимента

Было выбрано 1239 изображений людей, вместе с их зеркальным отражением. Всего 2478 положительных примеров и 12180 отрицательных, выбранные случайно из 1218 свободных от людей изображений. Далее для каждой комбинации параметров дескриптора, производился поиск сложных примеров, которые были ошибочно распознаны как положительные. Затем сложные пример добавлялись к отрицательным и на них уже обучался итоговый классификатор. Такой подход значительно улучшает точность: на 5%(miss rate) при 10^{-4} ложноположительных в одном окне. Повторные поиски не дают улучшения.

DET-кривые



Нормализация изображения в целом

Можно использовать гамма нормализацию. При значении гаммы меньше единицы слабо освещенные детали распознаются лучше. В статье используется гамма равная 0.5 , однако гамма коррекция изображения в целом даёт лишь малое изменения качества классификации.



Почему градиенты?

Градиенты позволяют зафиксировать изменение интенсивности в точке изображения. Гистограммы будут показывать доминирующее направление изменения картинки в данном месте. Такой подход позволяет хорошо распознавать объекты с четко выраженными границами.

Похожая модель описывает биологическое зрение, а именно сложные нейроны в зрительной коре. Эти сложные нейроны реагируют на градиенты конкретной ориентации и цвета, но позиция их может немного смещаться. То есть не используется информация о точном местоположении градиентов.

Вычисление градиентов в клетке для построения диаграммы

Интуитивное понятие градиента пикселя: мера того как сильно меняется картинка в этом месте.

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

Вектор градиента указывает на направление наибольшего роста интенсивности изображения

Градиенты вычисляются для одного канала или для серого изображения

Если изображение цветное то для каждого пикселя вычисляется градиент наибольшей магнитуды из трёх цветовых каналов

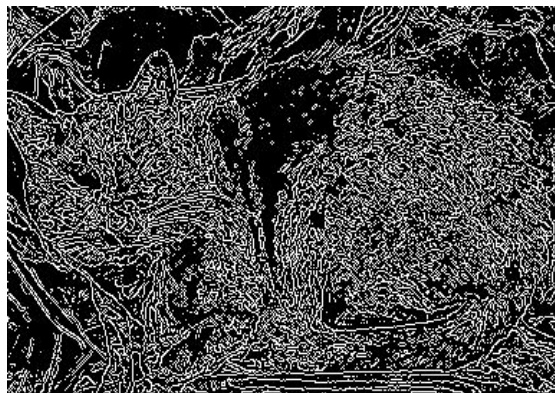
Гауссово размытие

Уменьшает детализированность изображения, что помогает в задаче распознавания границ. Оно усредняет изображение так, что четкие границы сохраняются.

Гауссова маска:

1	2	1
2	4	2
1	2	1





Способы вычисления градиентов

Точность детектора чувствительна к способу вычисления градиентов.

Оператор Собеля:

$$G_y = \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -2 & 0 & 2 \\ \hline -1 & 0 & 1 \\ \hline \end{array}$$

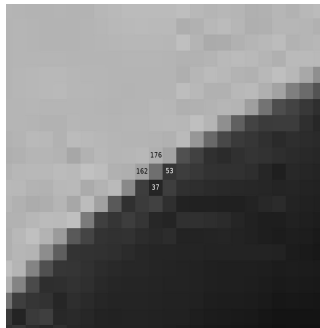
$$G_x = \begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

1-D-нецентрированная: $[-1, 1]$

1-D-центрированная: $[-1, 0, 1]$

2-D-диагональная: $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

Вычисление градиентов



На практике частные производные вычисляются как разность интенсивностей соседних пикселей.

$$\nabla f = \begin{pmatrix} 37 - 176 \\ 53 - 162 \end{pmatrix} = \begin{pmatrix} -139 \\ -109 \end{pmatrix}$$

$$|\nabla f| = \sqrt{139^2 + 109^2} = 177$$

$$\Theta = \arctan\left(\frac{-109}{-139}\right) = 38^\circ$$

Сглаживание лишь ухудшает точность, и простое вычисление градиента оказалось самым лучшим.

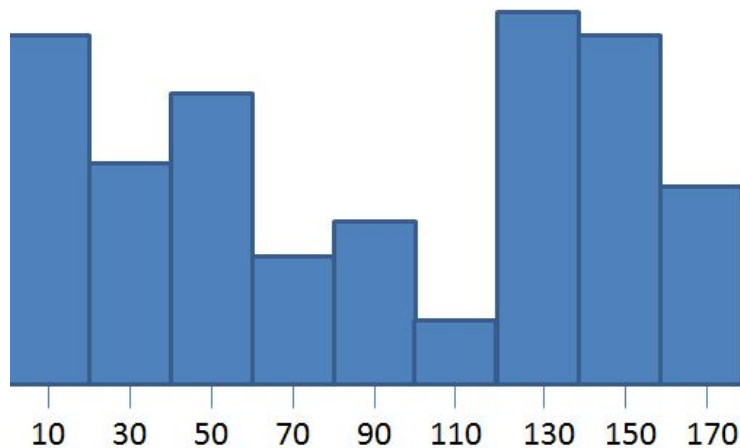
Составление гистограмм

Вклад каждого градиента это функция его длины. На практике используют саму длину.

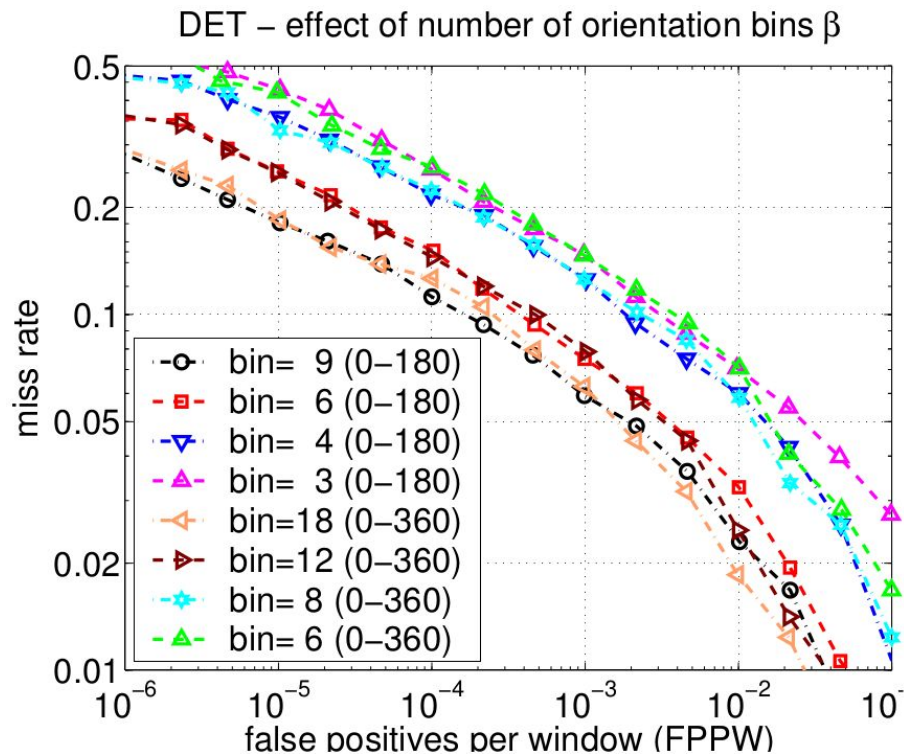
Значения необходимо интерполировать.

Если этого не делать точность ухудшается на

5% при 10^{-4} FPPW.



Число ориентаций

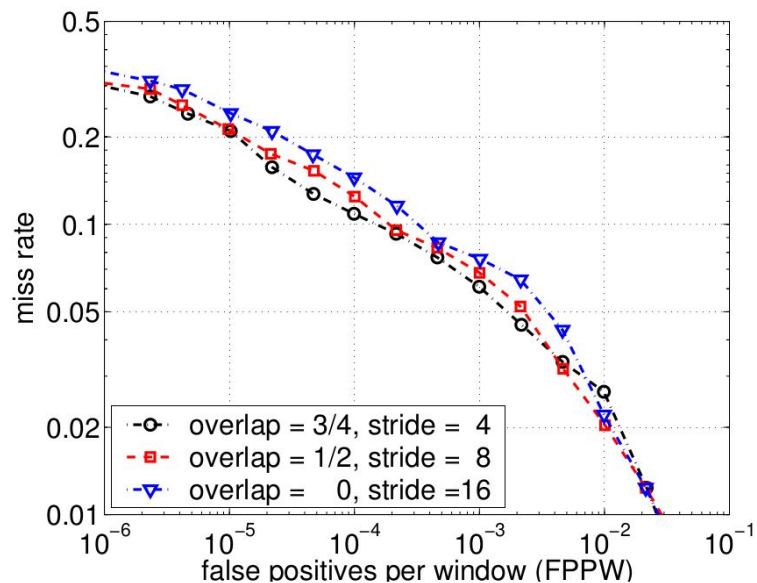


Локальная нормализация

Локально длина градиентов может изменяться из-за освещения и контраста переднего и заднего фона, поэтому эффективная контрастная нормализация оказывается необходимой для хорошей точности.

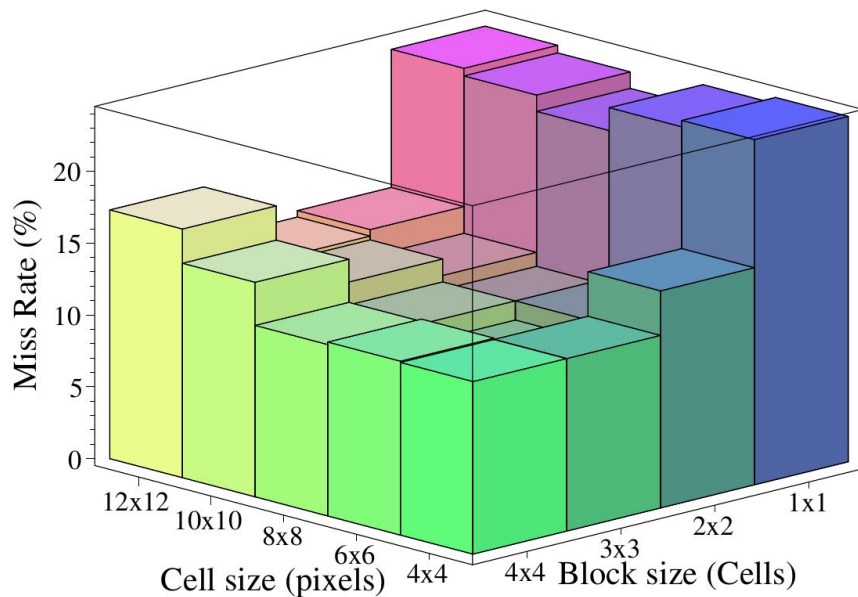
Способы нормализации

На практике блоки могут пересекаться и тогда каждая клетка вносит в финальный вектор несколько вкладов, каждый из которых нормализован в соответствии со своим блоком. Пересечение блоков сильно увеличивает точность.



R-Hog

Квадрат: блоки состоящие из $N \times N$ клеток, каждая из которых имеет форму квадрата $M \times M$ пикселей.



Виды блочной нормализации

Пусть \mathbf{v} не нормализованный вектор блока, $\|\mathbf{v}\|_k$ -k норма для $k=1,2$ и ε -небольшая константа.

a) L2-норма,

$$\mathbf{v} \rightarrow \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2^2 + \varepsilon^2}}$$

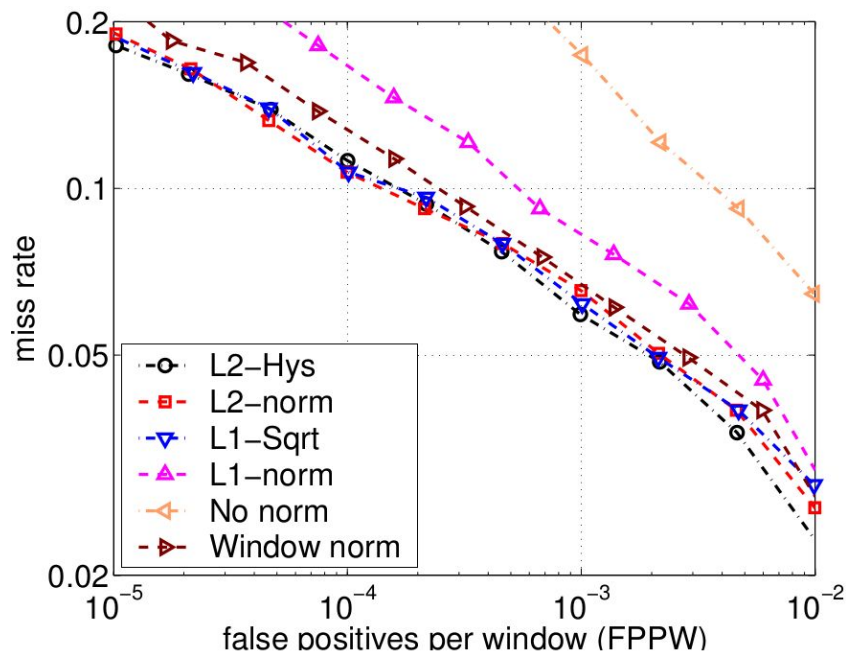
b) L2-Lowe.

c) L1-норма,

$$\mathbf{v} \rightarrow \frac{\mathbf{v}}{(\|\mathbf{v}\|_1 + \varepsilon)}$$

d) L1-sqrt,

$$\mathbf{v} \rightarrow \sqrt{\frac{\mathbf{v}}{(\|\mathbf{v}\|_1 + \varepsilon)}}$$



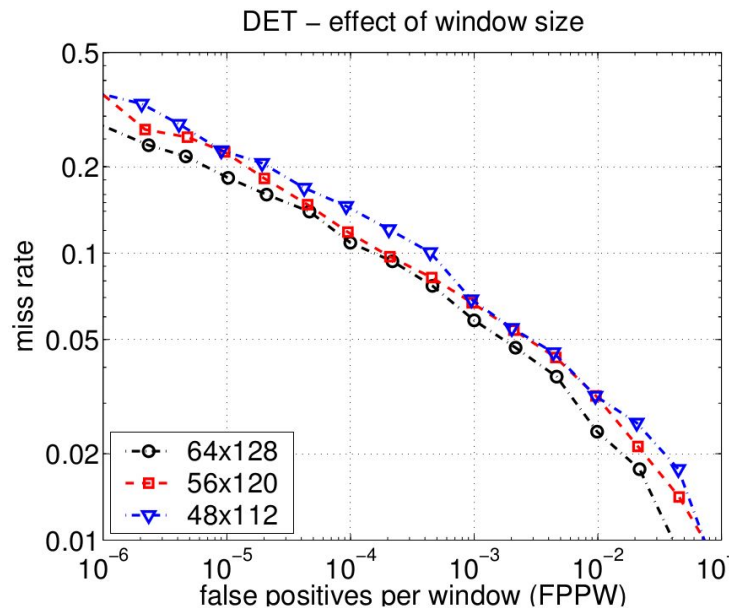
L2-lowe норма

Могут происходить нелинейные измерения изменения освещения изображения. Они могут сильно изменить относительную длину для некоторых градиентов, но они редко меняют их ориентацию. Поэтому имеет смысл уменьшить влияние градиентов с большой длиной, устанавливая пороговую длину 0,2. Таким образом сопоставление длин больших градиентов даёт меньший вклад в дескриптор и распределения градиентов по ориентациям имеет большее значение.

Размеры окна

Окно 64x128 даёт отступ от человека около 16 пикселей во все стороны.

Если уменьшать окно или увеличивать человека, то качество падает так как становится недостаточно контекста.



Beca SVM



Заключение

Ног декрипторы отлично подходят для распознавания людей на изображениях. По качеству работы их можно сравнивать со сверточным нейронными сетями.