Компьютерное зрение и обработка изображений
Лекция 13
Детекторы и дескрипторы. Продолжение

Я.М.Демяненко
demyana@sfedu.ru

Южный федеральный университет
Институт математики, механики и компьютерных наук

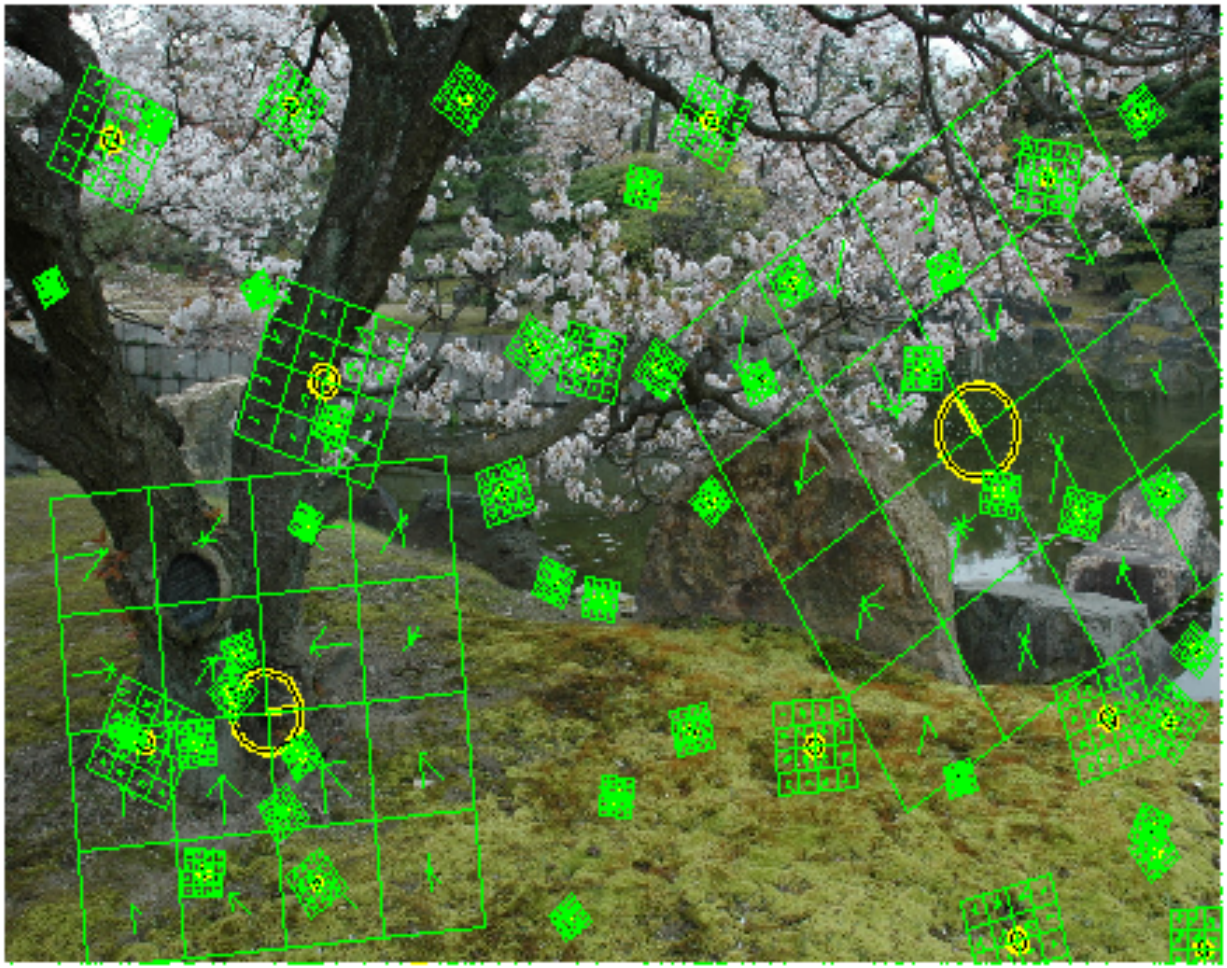
2019

Содержание

Содержание

SIFT(Scale Invariant Feature Transform)

- The algorithm was patented in the US by the University of British Columbia and published by David Lowe in 1999.



Основной момент

- Основным моментом в детектировании особых точек является построение пирамиды гауссианов (Gaussian) и разностей гауссианов (Difference of Gaussian, DoG).

Гауссиан и разность гауссианов

- Гауссианом (или изображением, размытым гауссовым фильтром) является изображение

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

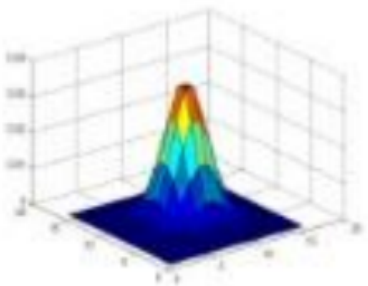
- L — значение гауссиана в точке с координатами (x,y), sigma — радиус размытия. G — гауссово ядро
- Разностью гауссианов называют изображение, полученное путем попиксельного вычитания одного гауссина исходного изображения из гауссиана с другим радиусом размытия.

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned}$$

Разность первого и второго гауссианов

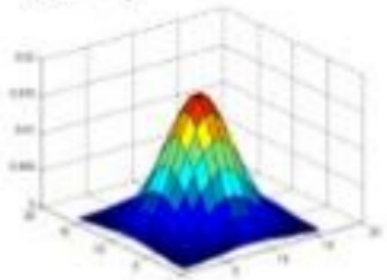
Оригинал

$\sigma_1 = 2$



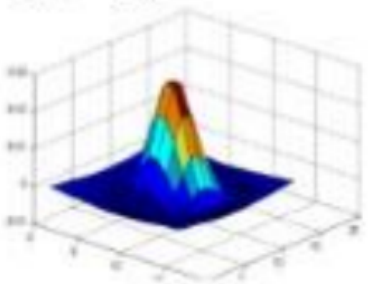
Первый гауссиан

$\sigma_2 = 4$



Второй гауссиан

$\sigma_1 - \sigma_2$

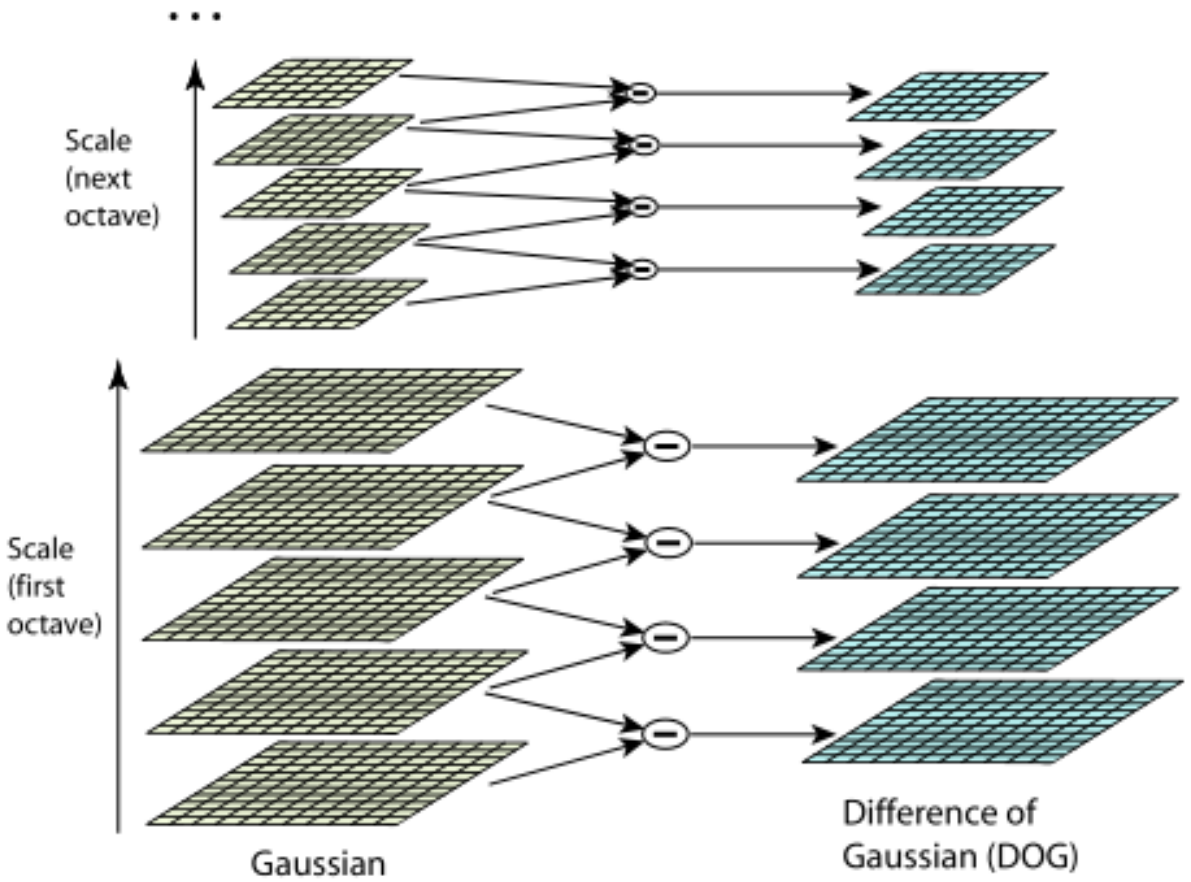


Разность гауссианов

Масштабируемое пространство изображения

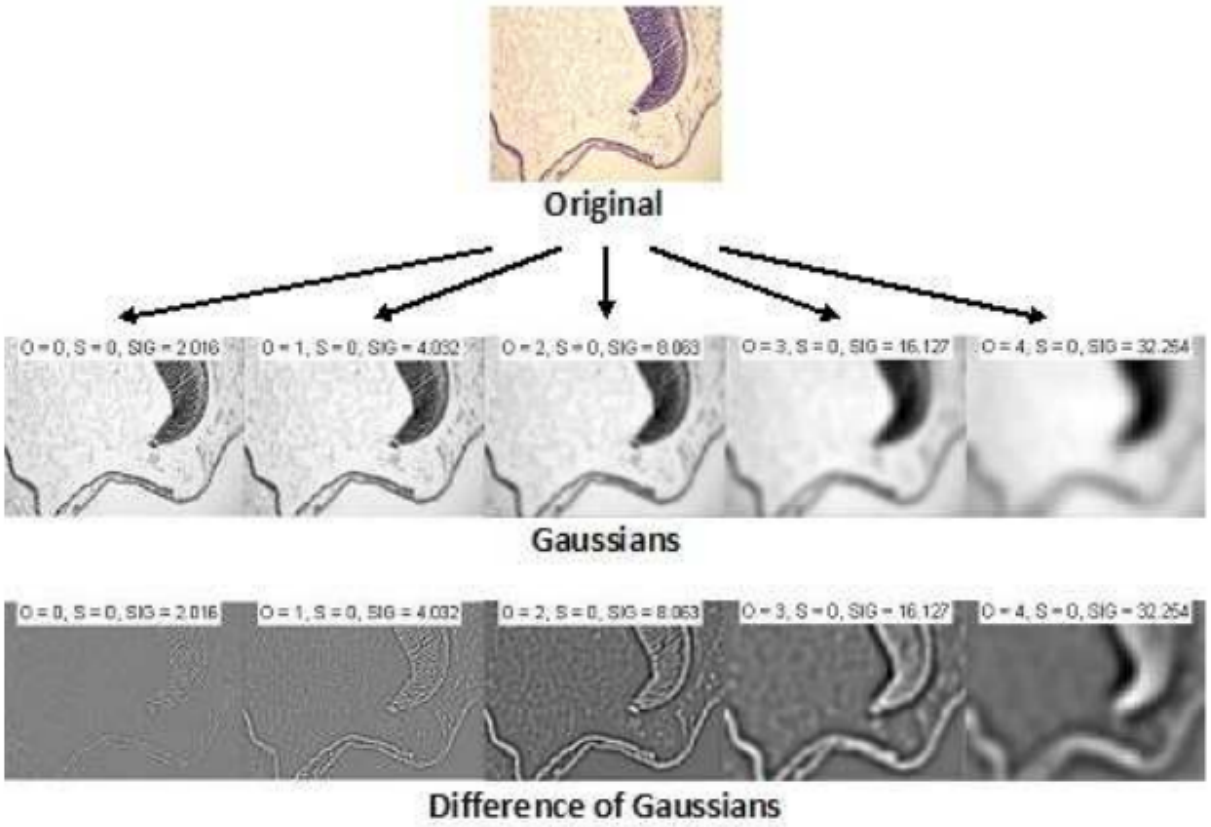
- Масштабируемым пространством изображения является набор всевозможных, сглаженных некоторым фильтром, версий исходного изображения.
- Доказано, что гауссово масштабируемое пространство является линейным, инвариантным относительно сдвигов, вращений, масштаба, не смещающим локальные экстремумы, и обладает свойством полугрупп.
- Различная степень размытия изображения гауссовым фильтром может быть принята за исходное изображение, взятое в некотором масштабе.

Пирамида гауссианов и пирамида DOG



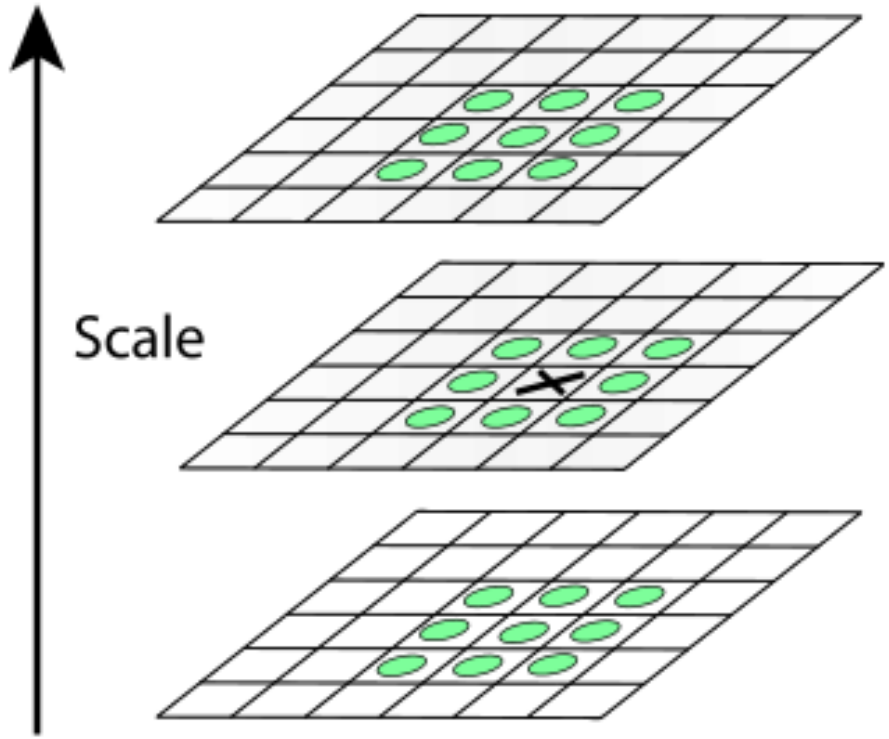
- Масштаб первого изображения следующей октавы равен масштабу изображения из предыдущей октавы с номером N.

Пример



- Пример гауссианов и разности гауссианов при различных значениях сигма.

Локальный экстремум – особая точка?



Будем считать точку особой, если она является локальным экстремумом разности гауссианов

Как уточнить?

- Это достигается с помощью аппроксимирования функции DoG многочленом Тейлора второго порядка, взятого в точке вычисленного экстремума.

$$D(\mathbf{x}) = D + \frac{\partial D^{\mathbf{T}}}{\partial \mathbf{x}}\,\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{T}}\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

- D — функция DoG, X = (x,y,sigma) — вектор смещения относительно точки разложения, первая производная DoG — градиент, вторая производная DoG — матрица Гессе.

Уточнение особых точек

- Экстремум многочлена Тейлора находится путем вычисления производной и приравнивания ее к нулю. В итоге получим смещение точки вычисленного экстремума, относительно точного

$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}$$

- Все производные вычисляются по формулам конечных разностей. В итоге получаем СЛАУ размерности 3x3, относительно компонент вектора $\hat{X}$.

Уточнить? А может выбросить?

- Если одна из компонент вектора $\hat{X}$ больше 0.5 шага сетки в этом направлении, то это означает, что на самом деле точка экстремума была вычислена неверно и нужно сдвинуться к соседней точке в направлении указанных компонент.
- Для соседней точки все повторяется заново. Если таким образом мы вышли за пределы октавы, то следует исключить данную точку из рассмотрения.

Отсечение по контрасту

- Когда положение точки экстремума вычислено, проверяется на малость само значение DoG в этой точке по формуле

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2}\frac{\partial D}{\partial \mathbf{x}}^T \hat{\mathbf{x}}$$

- Если эта проверка не проходит, то точка исключается, как точка с малым контрастом.

Исключение точек с малым контрастом

Отсечение на границе

- Если особая точка лежит на границе какого-то объекта, то такую точку можно исключить из рассмотрения.

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Пусть Tr(H) — след матрицы, а Det(H) — её определитель.

$$\mathrm{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta,$$

$$\mathrm{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta.$$

Пусть r — отношение большего изгиба к меньшему, $\alpha = r\beta$

тогда $$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r}$$

и точка рассматривается дальше, если $$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r+1)^2}{r}$$

Исключение точек на границе



Нахождение ориентации ключевой точки

- Направление ключевой точки вычисляется исходя из направлений градиентов точек, соседних с особой. Все вычисления градиентов производятся на изображении в пирамиде гауссианов, с масштабом наиболее близким к масштабу ключевой точки.
- Величина и направление градиента в точке (x,y) вычисляются по формулам

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$

m — величина градиента, theta — его направление

Окрестность ключевой точки

- Это окно, требуемое для свертки с гауссовым ядром, причем радиус размытия для этого ядра (sigma) равен 1.5 масштаба ключевой точки.
- Окно будет круглым и радиус окна определяется как 3*sigma

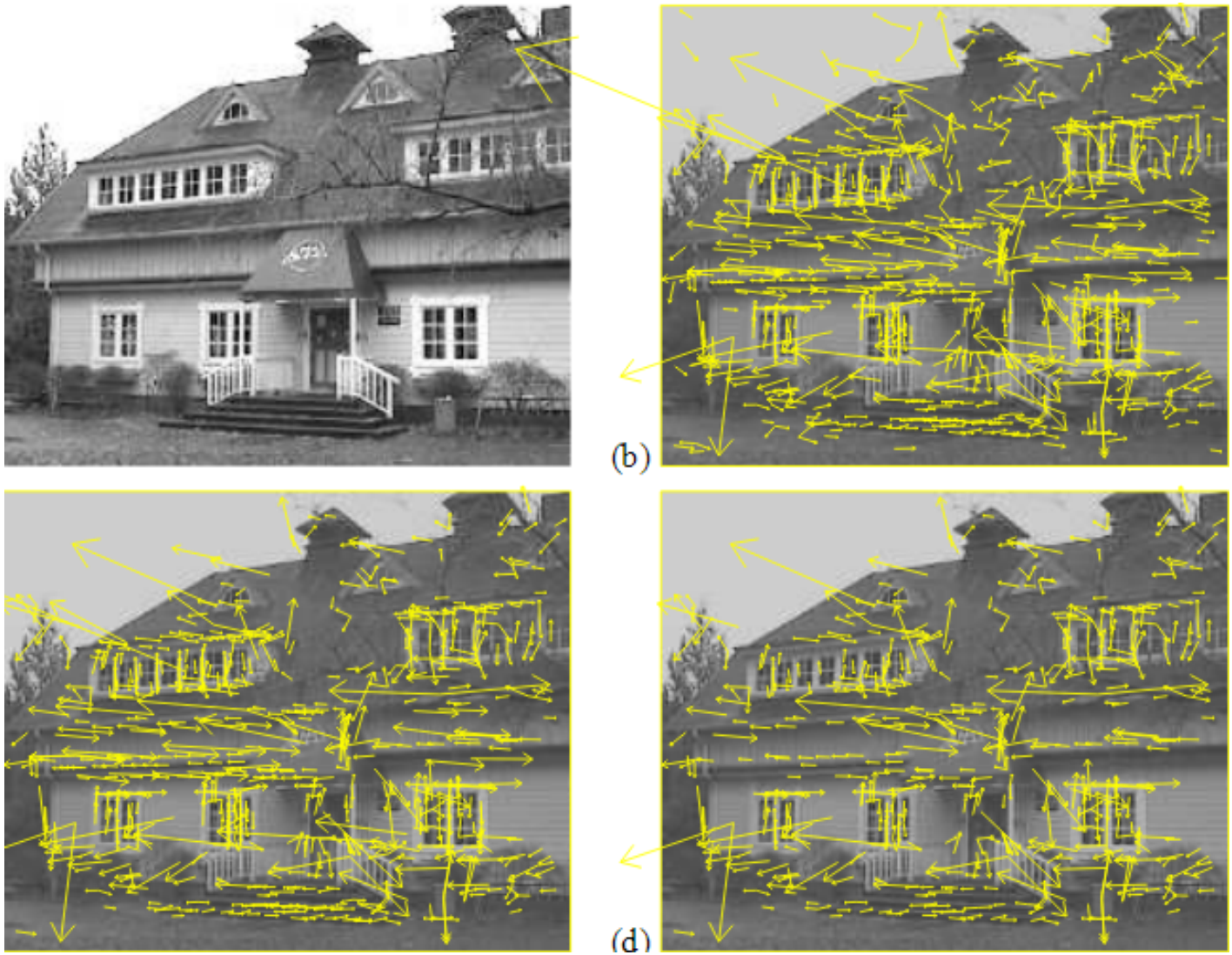
Гистограмма направлений

- Направление ключевой точки найдём из гистограммы направлений.
- Гистограмма состоит из 36 компонент, которые равномерно покрывают промежуток в 360 градусов, и формируется она следующим образом: каждая точка окна (x, y) вносит вклад, равный m*G(x, y, sigma), в ту компоненту гистограммы, которая покрывает промежуток, содержащий направление градиента theta(x, y).

Направление ключевой точки

- Направление ключевой точки лежит в промежутке, покрываемом максимальной компонентой гистограммы.
- Значения максимальной компоненты (max) и двух соседних с ней интерполируются параболой, и точка максимума этой параболы берётся в качестве направления ключевой точки.
- Если в гистограмме есть ещё компоненты с величинами не меньше 0.8*max, то они аналогично интерполируются и дополнительные направления приписываются ключевой точке.

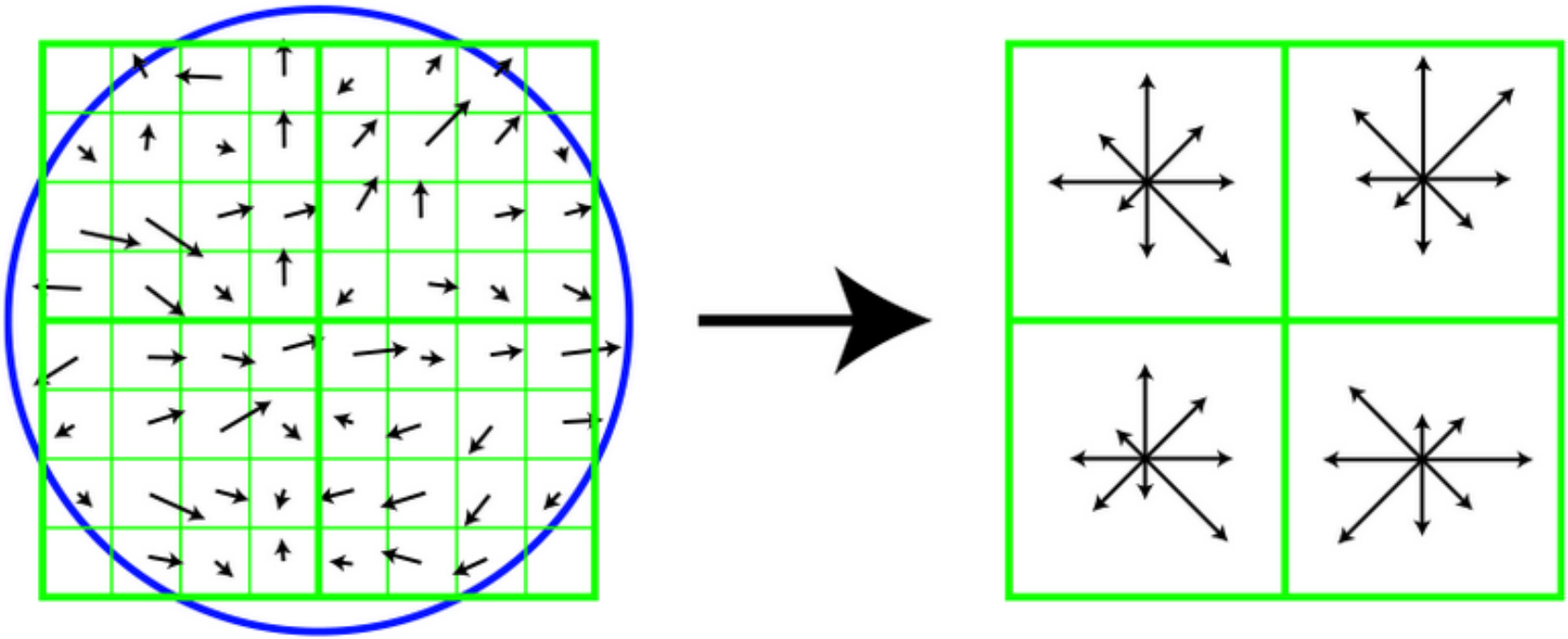
Этапы отбора точек



Построение дескрипторов

- Как и направление ключевой точки, дескриптор вычисляется на гауссиане, ближайшем по масштабу к ключевой точке, и исходя из градиентов в некотором окне ключевой точки.
- Перед вычислением дескриптора это окно поворачивают на угол направления ключевой точки, чем и достигается инвариантность относительно поворота.

Фрагмент с градиентами изображения и полученный на его основе дескриптор



Гистограммы в регионах

- Каждая гистограмма так же покрывает участок в 360 градусов, но делит его на 8 частей
- В качестве весового коэффициента берется значение гауссова ядра, общего для всего дескриптора

Построение дескрипторов

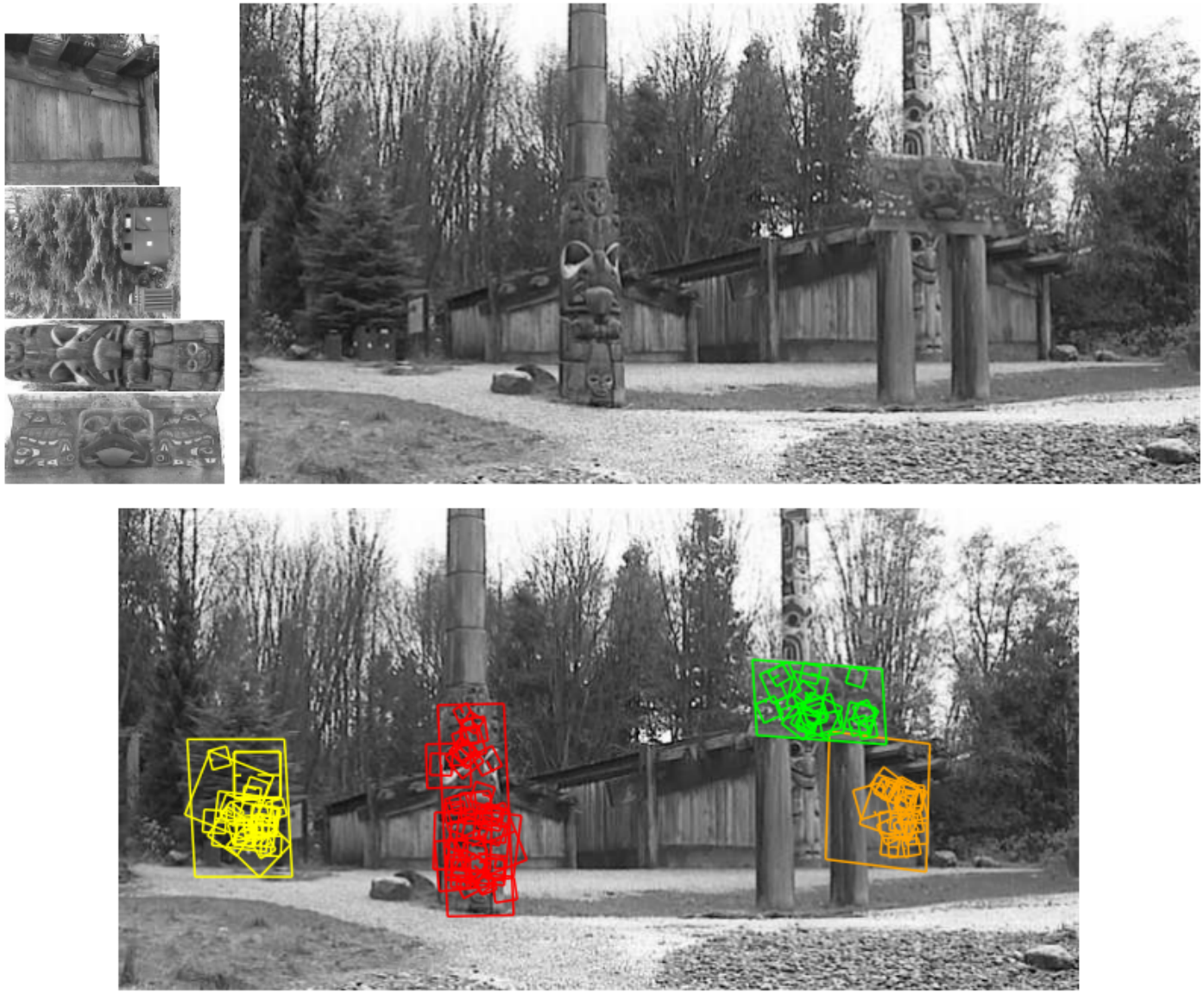
- Дескриптор ключевой точки состоит из всех полученных гистограмм.
- Размерность дескриптора на рисунке 32 компоненты (2x2x8), но на практике используются дескрипторы размерности 128 компонент (4x4x8).
- Полученный дескриптор нормализуется, после чего все его компоненты, значение которых больше 0.2, урезаются до значения 0.2 и затем дескриптор нормализуется ещё раз. В таком виде дескрипторы готовы к использованию.

Результаты


10
PACK
maxell
CD-R
700
MB
80 Minutes
12x


10
PACK
maxell
CD-R
700
MB
80 Minutes
12x

Результаты



Статьи

- David G. Lowe «Distinctive image features from scale-invariant keypoints»
- https://www.youtube.com/watch?v=NPcMS49V5hg
- Yu Meng «Implementing the Scale Invariant Feature Transform(SIFT) Method»

Содержание

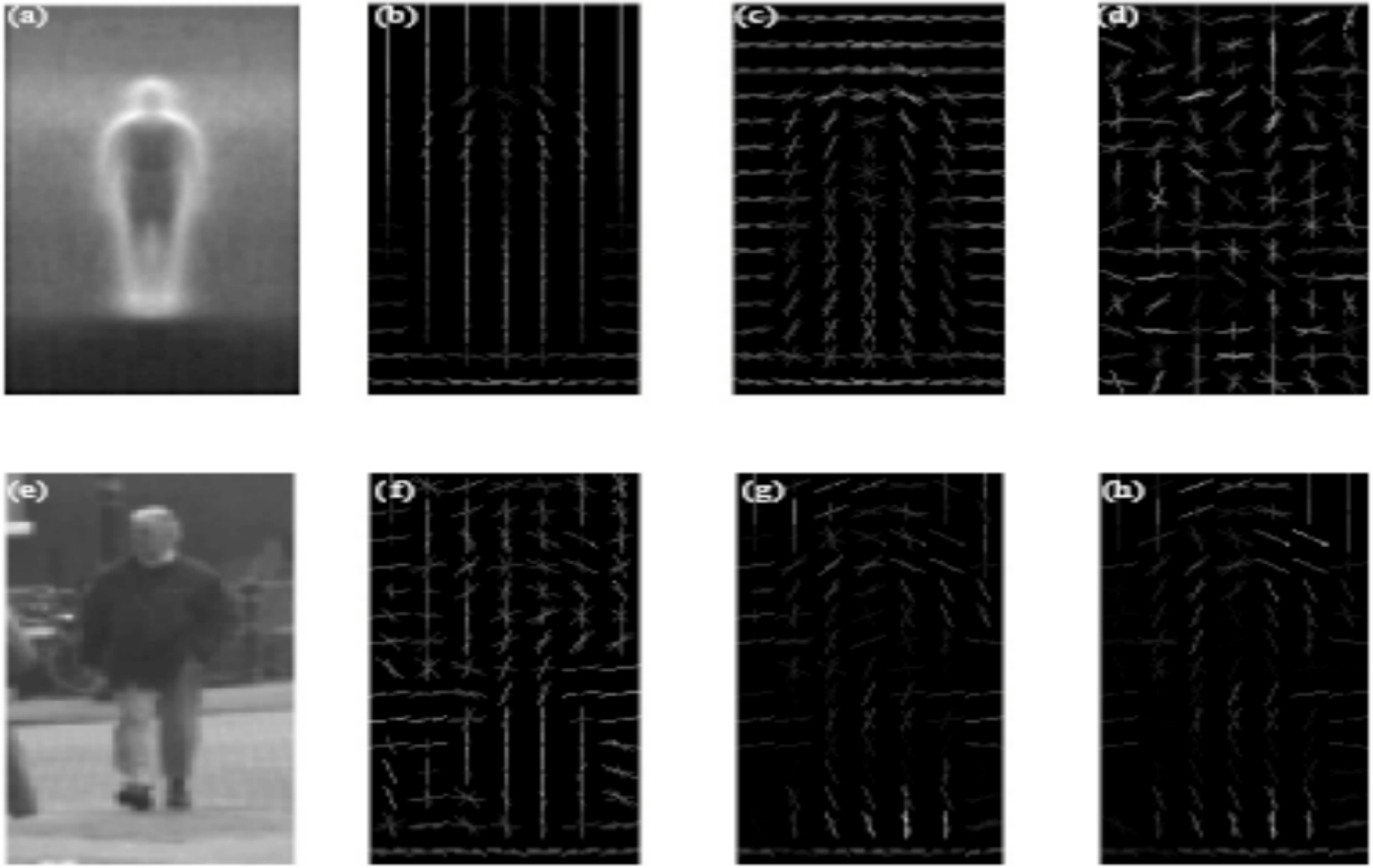
HOG

- Гистограмма направленных градиентов (англ. Histogram of Oriented Gradients, HOG) — дескрипторы особых точек
- HoG дескрипторы похожи на SIFT дескрипторы
- Но HoG дескрипторы вычисляются по всему изображению с единым масштабом без выравнивания ориентации
- SIFT дескрипторы вычисляются в ключевых точках изображения и разворачиваются для выравнивания ориентации

Первое описание HOG

- Навнит Далал и Билл Триггс, исследователи INRIA (Национальный исследовательский институт во Франции, работающий в области компьютерных наук, теории управления и прикладной математики.), впервые описали гистограмму направленных градиентов в своей работе на CVPR (Conference on Computer Vision and Pattern Recognition) в июне 2005 года.
- В этой работе они использовали алгоритм для нахождения пешеходов на статичных изображениях, хотя впоследствии расширили область применения до нахождения людей на видео, а также различных животных и машин на статичных изображениях.

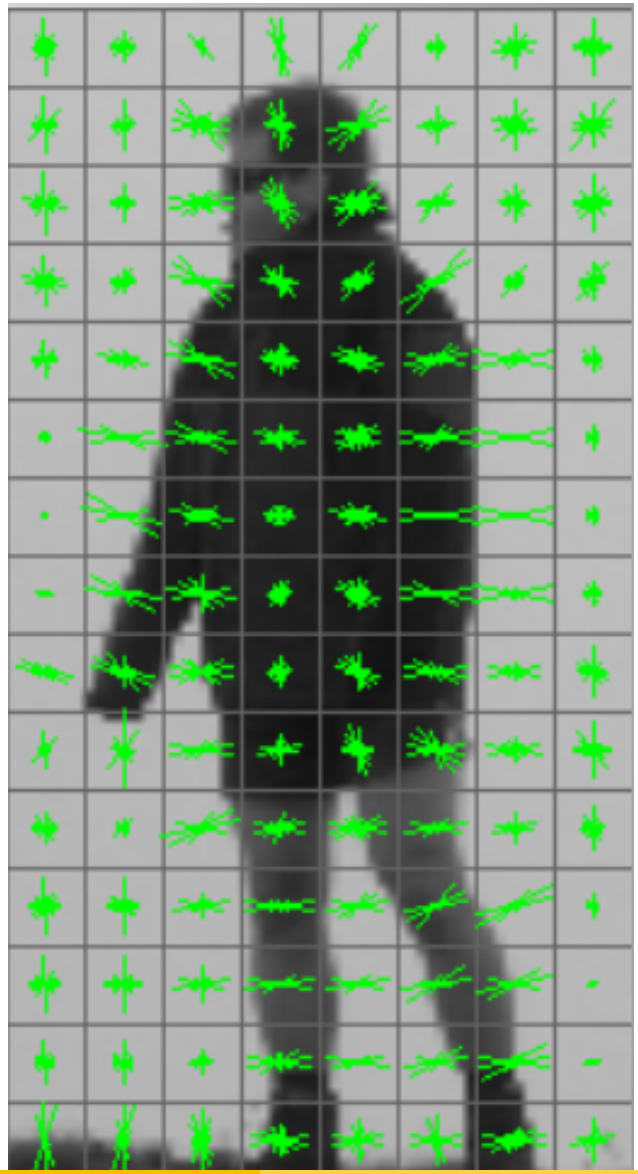
HOG для нахождения пешеходов



Идея подхода

- Форма и вид объектов на изображении могут хорошо описываться распределением относительных величин градиентов функции интенсивности, характеризующих направление границ объектов

Как вычисляются?



Основные шаги

- Разделение изображения на маленькие связные области — ячейки.
- Расчёт для каждой ячейки гистограммы направлений градиентов или направлений краев для пикселей, находящихся внутри ячейки. Комбинация этих гистограмм и является дескриптором.
- Для увеличения точности локальные гистограммы подвергаются нормализации по контрасту. С этой целью вычисляется мера интенсивности на большем фрагменте изображения, который называется блоком, и полученное значение используется для нормализации. Нормализованные дескрипторы обладают лучшей инвариантностью по отношению к освещению.

Плюсы и минусы

- Поскольку HOG работает локально, метод поддерживает инвариантность геометрических и фотометрических преобразований, за исключением ориентации объекта.
- Как обнаружили Далал и Триггс, грубое разбиение пространства, точное вычисление направлений и сильная локальная фотометрическая нормализация позволяют игнорировать движения пешеходов, если они поддерживают вертикальное положение тела.

Input image


Input image
Normalize gamma & colour
Compute gradients
Weighted vote into spatial & orientation cells
Contrast normalize over overlapping spatial blocks
Collect HOG's over detection window
Linear SVM
Person / non-person classification

Нормализация цвета

- Первым шагом вычислений во многих детекторах особых точек является нормализация цвета и гамма-коррекция.
- Далал и Триггс установили, что для дескриптора HOG этот шаг можно опустить, поскольку последующая нормализация даст тот же результат.
- Поэтому на первом шаге рассчитываются значения градиентов

Эксперименты с вычислением градиента

- Самым распространенным методом является применение одномерной дифференцирующей маски в горизонтальном и(или) вертикальном направлении

$$[-1, 0, 1] \text{ и } [-1, 0, 1]^T$$

- Далал и Триггс использовали более сложные маски, такие как Собел 3x3 или диагональные маски, но эти маски показали более низкую производительность для данной задачи.
- Они также экспериментировали с размытием по Гауссу перед применением дифференцирующей маски, но также обнаружили, что пропуск этого шага увеличивает быстродействие без заметной потери качества

Вычисление градиента

-1	0	1

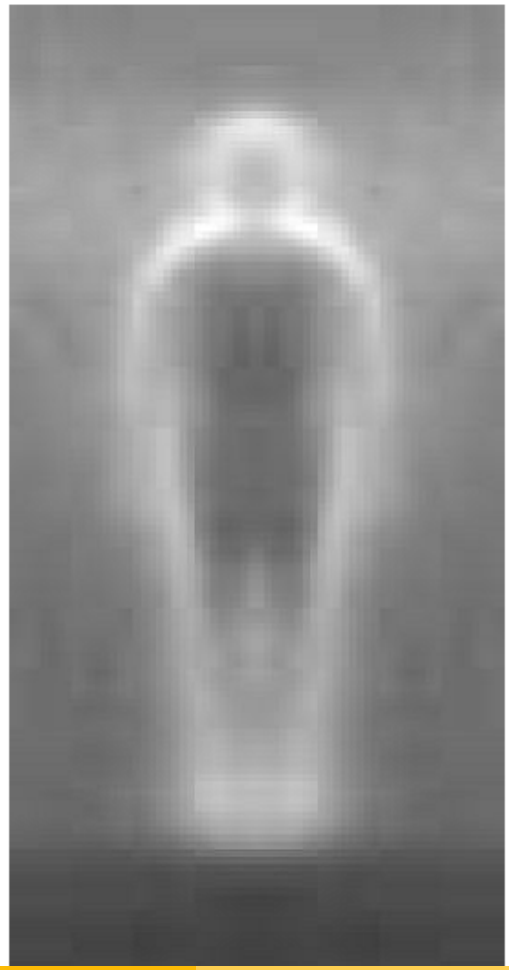
centered

-1	1

uncentered

1	-8	0	8	-1

cubic-corrected



0	1
-1	0

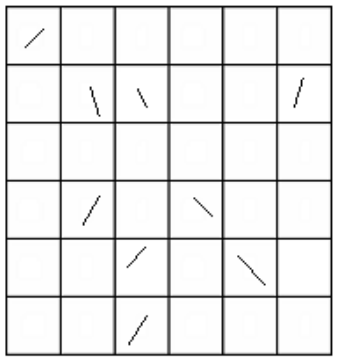
diagonal

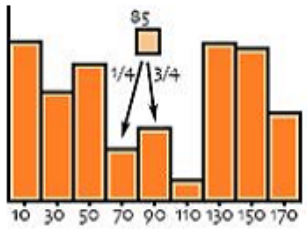
-1	0	1
-2	0	2
-1	0	1

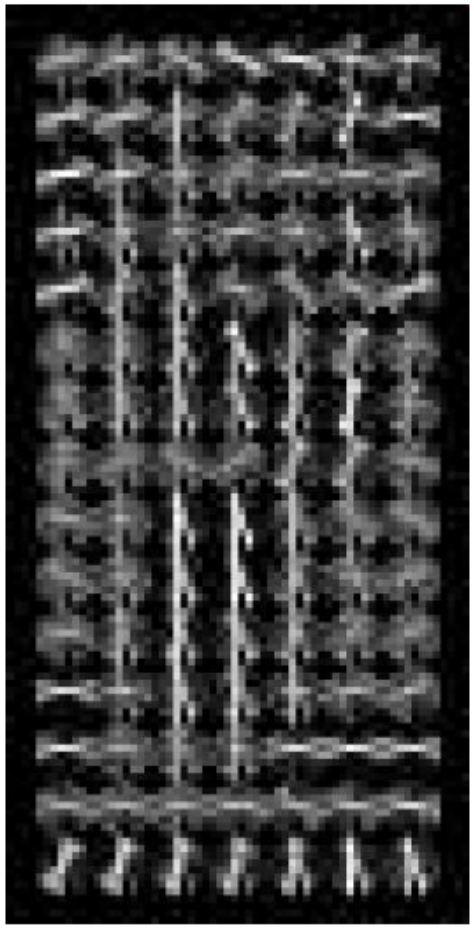
Sobel

Гистограммы ячеек

- Разбиваем картинку на блоки 8x8;
- Смотрим на градиенты в пикселях;
- Считаем локальную гистограмму градиентов для каждой ячейки.







Группировка направлений

- Вычисляются гистограммы ячеек
- Каждый пиксел в ячейке участвует во взвешенном голосовании для каналов гистограммы направлений, основанном на значении градиентов.
- Ячейки могут быть прямоугольной или круглой формы
- Каналы гистограммы равномерно распределяются от 0 до 180 или же от 0 до 360 градусов, в зависимости от того, вычисляется «знаковый» или «беззнаковый градиент».

Группировка направлений

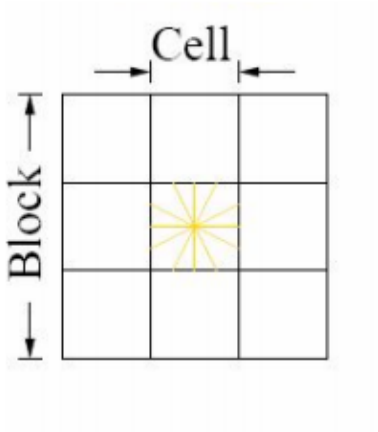
- Далал и Триггс обнаружили, что беззнаковый градиент совместно с девятью каналами гистограммы дает лучшие результаты при распознавании людей.
- При распределении весов в голосовании вес пикселя может задаваться либо абсолютным значением градиента, либо некоторой функцией от него; в реальных тестах абсолютное значение градиента дает лучшие результаты.
- Другими возможными вариантами могут быть квадратный корень, квадрат или урезанное абсолютное значение градиента

Блоки дескрипторов

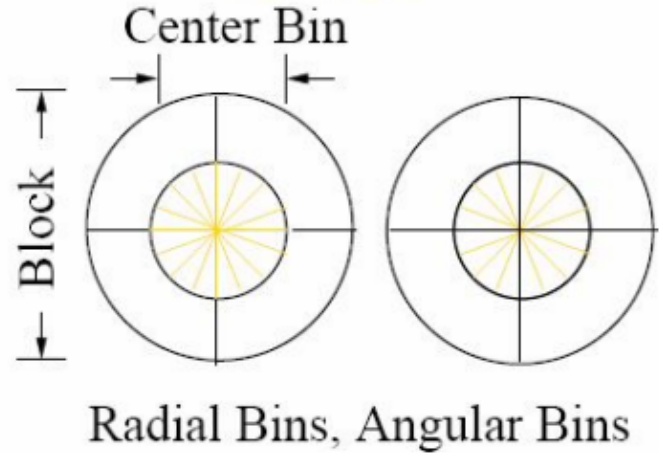
- Ячейки нужно сгруппировать в более крупные связные блоки
- Вычисляются меры интенсивности этих блоков
- Градиенты локально нормируются
- Дескриптор HOG, таким образом, является вектором компонент нормированных гистограмм ячеек из всех областей блока.
- Как правило, блоки перекрываются, то есть каждая ячейка входит более чем в один конечный дескриптор.
- Используются две основные геометрии блока: прямоугольные R-HOG и круглые C-HOG.

Блоки дескрипторов

- R-HOG
  - точный размер
  - квадратный блок



- C-HOG
  - цельная центральная ячейка
  - разделенная на сектора



Блоки R-HOG

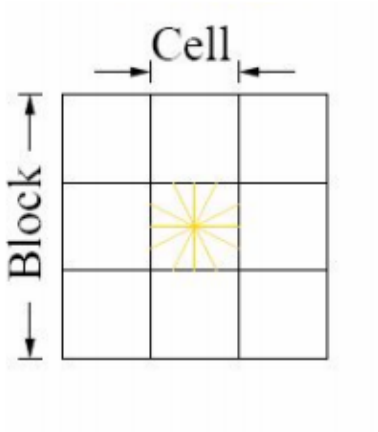
- Блоки R-HOG обычно являются квадратными сетками, характеризующимися тремя параметрами: количеством ячеек на блок, количеством пикселов на ячейку и количеством каналов на гистограмму ячейки.
- В эксперименте Далала и Триггса оптимальными параметрами являются блоки 16x16, ячейки 8x8 и 9 каналов на гистограмму.
- Более того, они обнаружили, что можно слегка повысить скорость вычислений, применяя гауссов фильтр внутри каждого блока до процедуры голосования, что, в свою очередь, снижает вес пикселей на границах блоков.

Блоки R-HOG и SIFT-дескрипторы

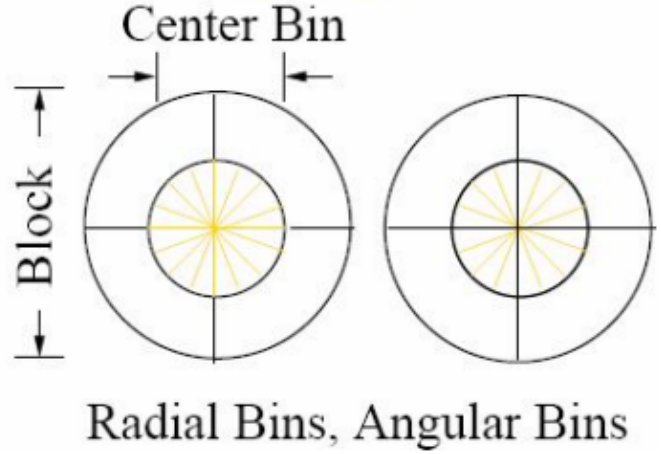
- Блоки R-HOG оказываются очень похожими на SIFT-дескрипторы
- Однако, блоки R-HOG вычисляются на плотных сетках фиксированного масштаба без фиксированного направления
- SIFT-дескрипторы вычисляются в разреженных, не чувствительных к масштабу ключевых точках изображения и используют поворот для выравнивания направления.
- Кроме того, для кодирования информации о форме объектов блоки R-HOG используются совместно, в то время как SIFT-дескрипторы используются по отдельности.

Блоки дескрипторов

- R-HOG
  - точный размер
  - квадратный блок



- C-HOG
  - цельная центральная ячейка
  - разделенная на сектора



Блоки C-HOG

- Блоки C-HOG имеют 2 разновидности: с цельной центральной ячейкой и разделенной на сектора.
- Эти блоки могут быть описаны 4 параметрами: количество секторов и колец, радиус центрального кольца и коэффициент расширения для радиусов остальных колец.
- Далал и Триггс обнаружили, что обе разновидности показали одинаковый результат, и разделение на 2 кольца и 4 сектора с радиусом 4 пиксела и коэффициентом расширения 2 дало лучший результат в их эксперименте.
- Кроме того, гауссово взвешивание не дало никаких улучшений при использовании блоков C-HOG.
- Эти блоки похожи на контексты формы, но имеют важное отличие: блоки C-HOG содержат ячейки с несколькими каналами направлений, в то время как контексты формы используют только наличие одного края.

Нормализация блоков

$$L1 - norm : v \longrightarrow v/(||v||_1 + \epsilon)$$

$$L2 - norm : v \longrightarrow v/\sqrt{||v||_2^2 + \epsilon^2}$$

$$L1 - sqrt : v \longrightarrow \sqrt{v/(||v||_1 + \epsilon)}$$

$L2 - hys$: L2-norm, plus clipping at .2 and renomalizing

Нормализация блоков

- Далал и Триггс исследовали четыре метода нормализации блоков.
- Пусть v — ненормированный вектор, содержащий все гистограммы данного блока, e — некая малая константа (точное значение не так важно). Тогда нормировочный множитель можно получить одним из следующих способов:

$$L1 - norm : v \longrightarrow v/(||v||_1 + \epsilon)$$

$$L2 - norm : v \longrightarrow v/\sqrt{||v||_2^2 + \epsilon^2}$$

$$L1 - sqrt : v \longrightarrow \sqrt{v/(||v||_1 + \epsilon)}$$

$$L2 - hys : \text{L2-norm, plus clipping at .2 and renomalizing}$$

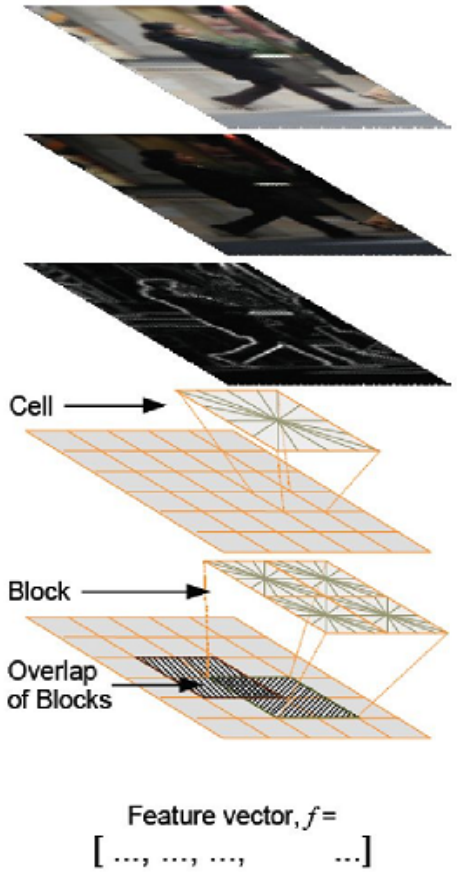
- L1-норма дает менее надежные результаты, чем остальные три, которые работают приблизительно одинаково хорошо, однако все четыре метода значительно улучшают результаты по сравнению с ненормализованными

Размер вектора-признака

Размер вектора-признака:

- 16 x 8 (тайлинг)
- 8 (ориентаций)
- 4 (блоки с перекрытиями)

Итого, 4096.



SVM-классификатор

- Конечным шагом в распознавании объектов с использованием HOG является классификация дескрипторов при помощи системы обучения с учителем.
- Далал и Триггс использовали метод опорных векторов (SVM, Support Vector Machine)

Итого

- Считаем градиенты в каждом пикселе картинки;
- Считаем гистограмму градиентов в каждой ячейке;
- Группируем ячейки в блоки (с перекрытиями);
- Нормализуем каждый блок;
- Склеиваем гистограммы из всех блоков в один вектор;
- Тренеруем линейный SVM для классификации.

Итого

- Фильтры [-1 0 1] и [-1 0 1]T для вычисления градиентов достаточно хороши для данной задачи;
- Углы от 0 до 180 квантуются по 9;
- Размер окна - 64x128, ячейки - 8x8, блоков - 16x16;
- При группировке ячеек в блоки, каждая ячейка учитывается 4 раза.

Тестирование База MIT

- База данных пешеходов Массачуссетского технологического института содержит обучающую выборку из 509 изображений и тестовую выборку из 200 изображений.
- Набор содержит изображения людей только спереди или сзади, позы на изображениях почти не отличаются.
- Эта база данных широко известна и используется в других исследованиях, найти её можно по ссылке
- http://cbcl.mit.edu/cbcl/software-datasets/PedestrianData.html

Тестирование База INRIA

- Второй набор данных был специально создан Далалом и Триггсом для их эксперимента, поскольку на наборе MIT дескрипторы HOG показали почти совершенные результаты.
- Этот набор данных, известный как INRIA, содержит 1805 изображений людей.
- Набор содержит изображения людей в широком разнообразии поз, включает в себя изображения с трудным фоном (например, на фоне толпы), и является гораздо более сложным для распознавания, чем набор MIT.
- База данных INRIA в настоящий момент доступна по адресу http://lear.inrialpes.fr/data

Тестирование

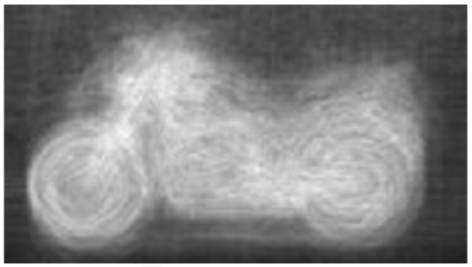
- В оригинальном эксперименте по обнаружению людей, Далал и Триггс сравнивали дескрипторы R-HOG и C-HOG с обобщенными вейвлетами Хаара и контекстами формы

Дескриптор	Набор изображений	Доля пропущенных изображений	Доля ошибок первого рода
HOG	MIT	≈0	10^{-4}
HOG	INRIA	0.1	10^{-4}
Обобщенные вейвлеты Хаара	MIT	0.01	10^{-4}
Обобщенные вейвлеты Хаара	INRIA	0.3	10^{-4}
PCA-SIFT, контексты формы	MIT	0.1	10^{-4}
PCA-SIFT, контексты формы	INRIA	0.5	10^{-4}

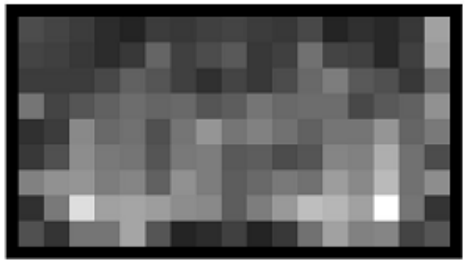
Дальнейшее развитие

- В рамках семинара Pascal Visual Object Classes в 2006 году, Далал и Триггс представили результаты применения HOG-дескрипторов к поиску на изображениях не только людей, но и машин, автобусов, велосипедов, собак, кошек и коров, а также оптимальные параметры для формирования и нормализации блоков в каждом случае.

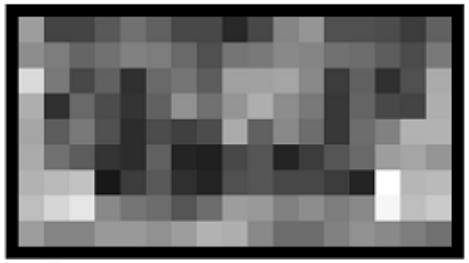
Мотоциклы



Градиенты



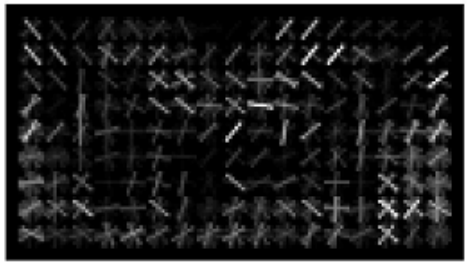
Взвешенные + веса



Взвешенные - веса

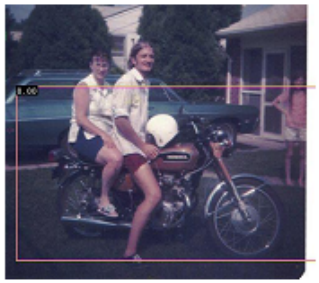
Окно

Доминирующие + ориентации

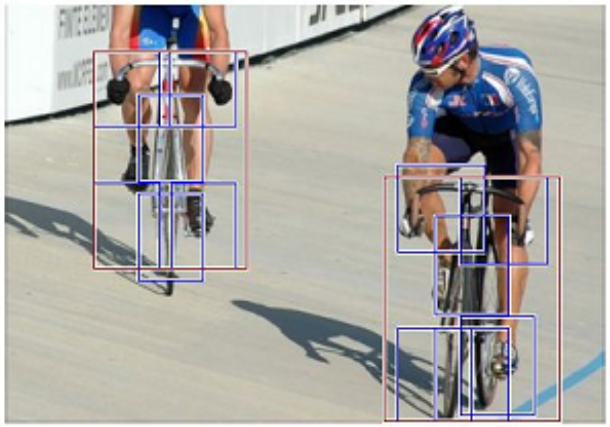


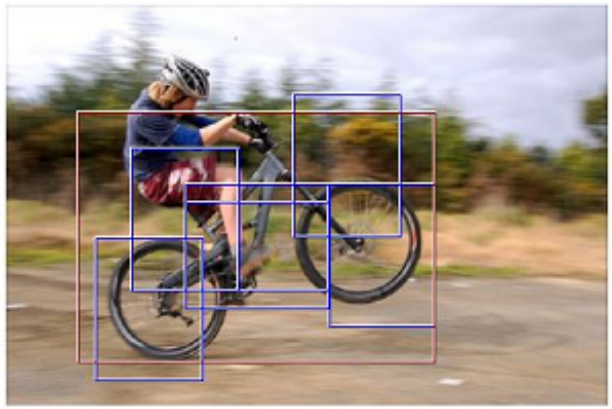
Доминирующие - ориентации

Примеры работы

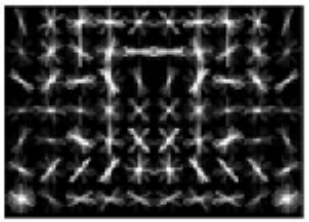


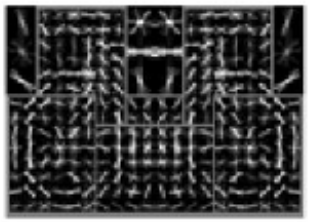
Модель велосипеда

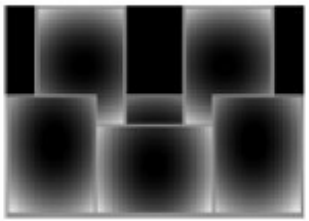




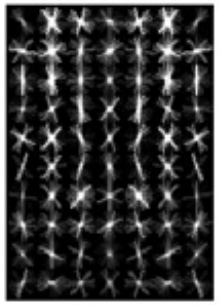
Вид сбоку

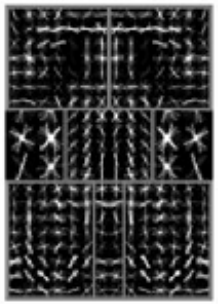


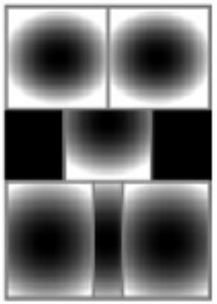




Вид спереди

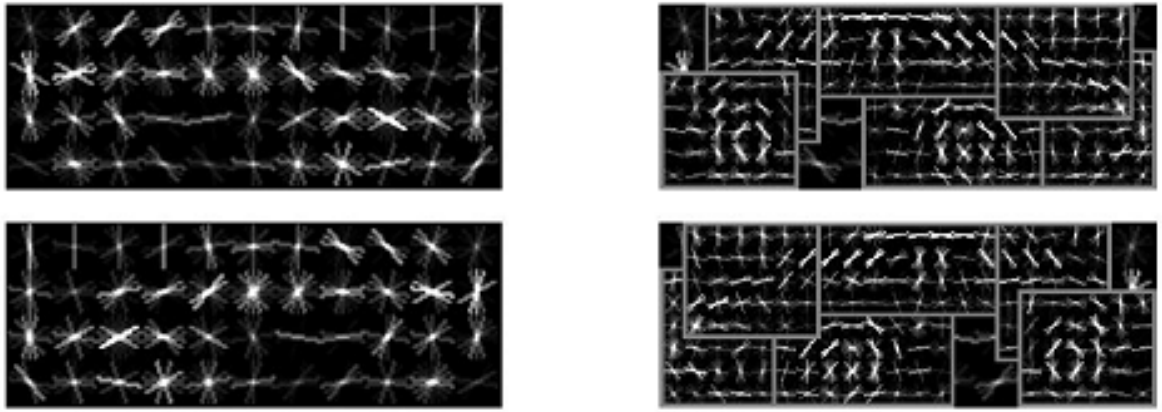




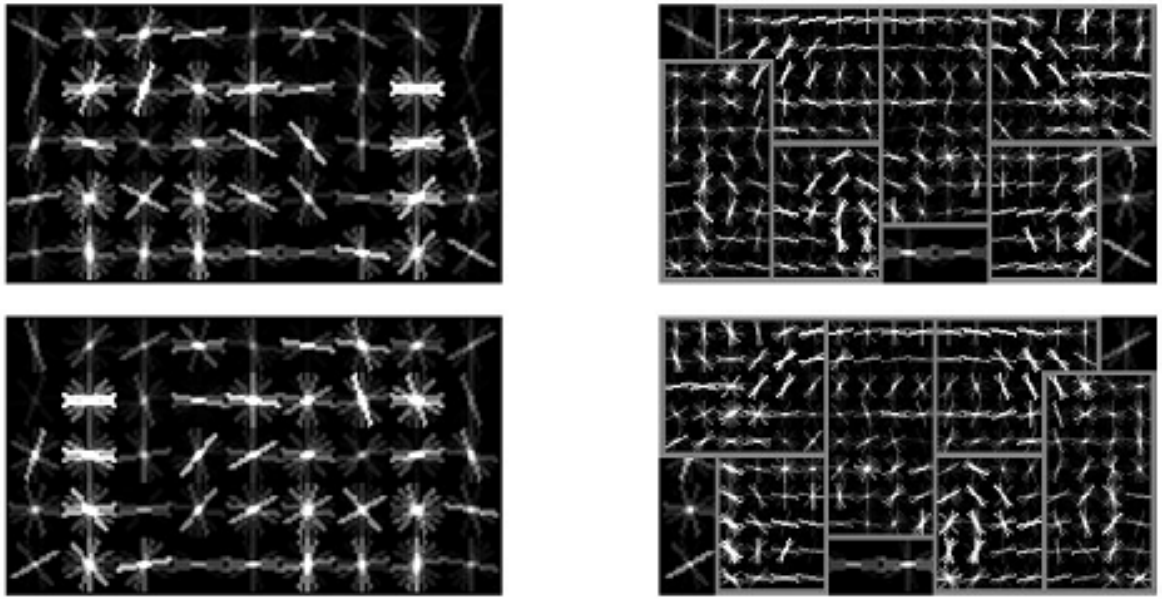


Модель автомобиля

Вид сбоку



Вид спереди



Содержание

Ключевые точки

- Использует FAST для нахождения ключевых точек – модификация FAST-9
- После выявления потенциальных ключевых точек используется угловой детектор Харриса для их уточнения.
- Чтоб получить N ключевых точек, сначала используется низкий порог для того, чтоб получить больше N точек, затем они упорядочиваются при помощи метрики Харриса и выбираются первые N точек.

Ключевые точки

- Использует FAST для нахождения ключевых точек – модификация FAST-9
- После выявления потенциальных ключевых точек используется угловой детектор Харриса для их уточнения.
- Чтоб получить N ключевых точек, сначала используется низкий порог для того, чтоб получить больше N точек, затем они упорядочиваются при помощи метрики Харриса и выбираются первые N точек.

Дескриптор

- Для построения дескриптора полученных точек используется модификация BRIEF, инвариантная к повороту за счет дополнительных преобразований

ORB

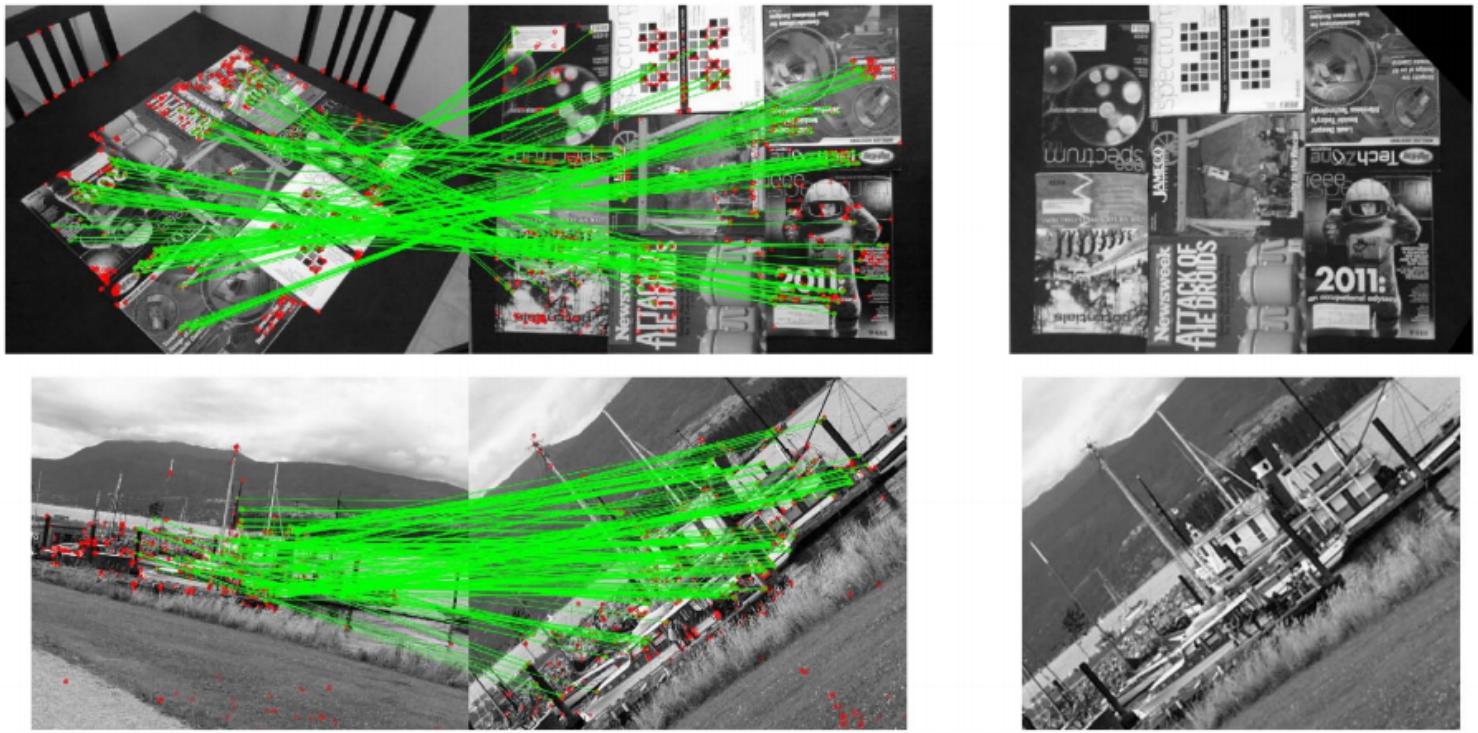


Figure 9. Real world data of a table full of magazines and an outdoor scene. The images in the first column are matched to those in the second. The last column is the resulting warp of the first onto the second.

Сравнение

	inlier %	N points
Magazines		
ORB	36.180	548.50
SURF	38.305	513.55
SIFT	34.010	584.15
Boat		
ORB	45.8	789
SURF	28.6	795
SIFT	30.2	714

Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski "ORB: an effcient alternative to SIFT or SURF"