НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Донейросетевые подходы к работе с изображениями

Научно-исследовательский семинар
Факультет компьютерных наук

Морозов Никита
БПМИ-182

12.10.2020

Введение

У людей возникает желание делать много всего интересного и полезного с изображениями:

Введение

У людей возникает желание делать много всего интересного и полезного с изображениями:

- распознавание объектов на изображениях по какой-то базе
- распознавание жестов
- слежение за движением объекта по нескольким снимкам
- реконструкция трехмерной модели объекта по его двумерным проекциям

И многое другое.

Введение

У людей возникает желание делать много всего интересного и полезного с изображениями:

- распознавание объектов на изображениях по какой-то базе
- распознавание жестов
- слежение за движением объекта по нескольким снимкам
- реконструкция трехмерной модели объекта по его двумерным проекциям

И многое другое.

Идея - сопоставлять изображениям или их частям наборы признаков, в каком-то смысле ”очень хорошо” их описывающие

HOG

HOG

Основные шаги:

- Найти градиенты во всех точках небольшой области (например клетки 8 на 8 пикселей)
- Построить гистограмму распределения градиентов
- В области побольше (например 128 на 64 пикселя) сложить гистограммы некоторых клеток в большой вектор
- Нормировать вектор

Градиенты

$$G_x(r, c) = I(r, c+1) - I(r, c-1)$$

$$G_y(r, c) = I(r-1, c) - I(r+1, c)$$

где $I(x, y)$ - яркость пикселя с координатами x, y

Градиенты

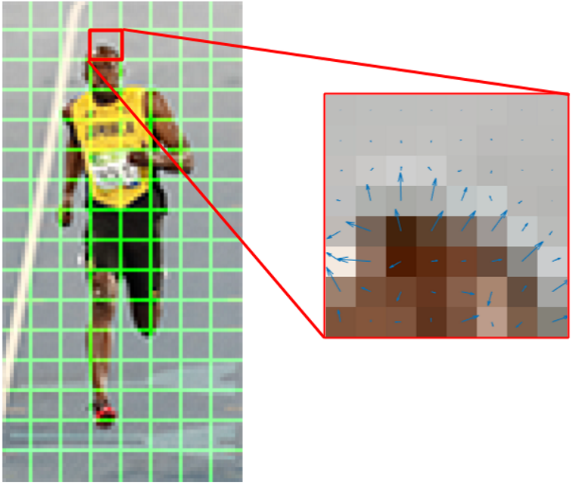
$$G_x(r,c) = I(r,c+1) - I(r,c-1)$$
$$G_y(r,c) = I(r-1,c) - I(r+1,c)$$

где $I(x,y)$ - яркость пикселя с координатами x, y

Переведем в полярные координаты:

$$\mu = \sqrt{G_x^2 + G_y^2}$$

$$\theta = \tan^{-1} \frac{G_y(x,y)}{G_x(x,y)}$$



2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

Gradient Magnitude

80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Gradient Direction

Гистограмма

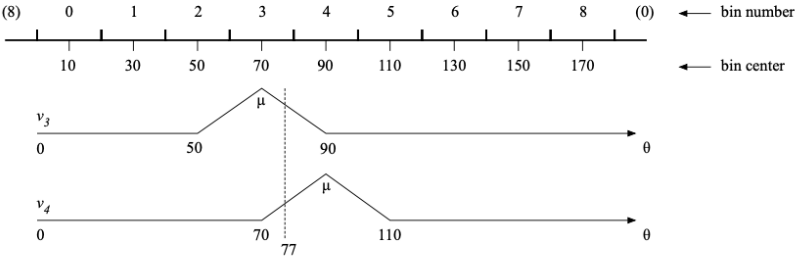
- Разобьем возможные значения угла на *B* ячеек

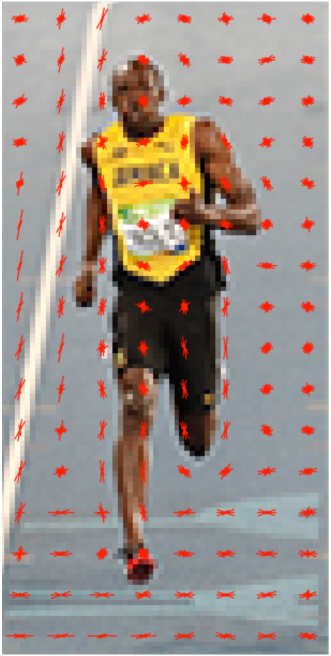
Гистограмма

- Разобьем возможные значения угла на B ячеек
- Каждый вектор дает вклад равный его длине

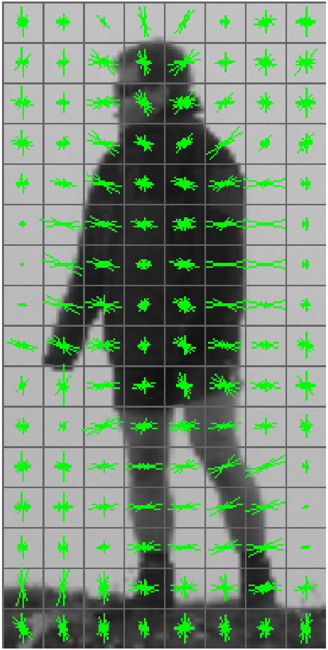
Гистограмма

- Разобьем возможные значения угла на *B* ячеек
- Каждый вектор дает вклад равный его длине
- Вклад разбивается между двумя ближайшими к значению угла ячейками пропорционально расстояниям до их центров





Визуализация гистограмм направленных градиентов в клетках изображения. Здесь угол градиентов считается от 0 до π, поэтому в каждой клетке гистограмма рисуется симметричной.



Дескрипторы блоков и нормализация

- Сгруппировать клетки в блоки 2 на 2 клетки

Дескрипторы блоков и нормализация

- Сгруппировать клетки в блоки 2 на 2 клетки
- Сложить четыре гистограммы в единый дескриптор блока **b** и нормировать его

$$\mathbf{b} \leftarrow \frac{\mathbf{b}}{\sqrt{\|\mathbf{b}\|^2 + \epsilon}}$$

Дескрипторы блоков и нормализация

- Сгруппировать клетки в блоки 2 на 2 клетки
- Сложить четыре гистограммы в единый дескриптор блока **b** и нормировать его

$$\mathbf{b} \leftarrow \frac{\mathbf{b}}{\sqrt{\|\mathbf{b}\|^2 + \epsilon}}$$

- Сложить дескритпоры блоков в единый HOG дескриптор и нормировать его

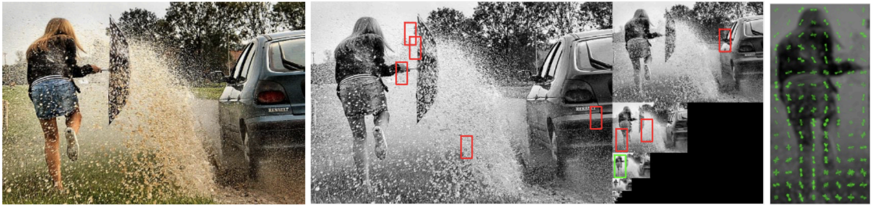
$$\mathbf{h} \leftarrow \frac{\mathbf{h}}{\sqrt{\|\mathbf{h}\|^2 + \epsilon}}$$

$$\mathbf{h}_i \leftarrow \min\left(\mathbf{h}_i, \tau\right)$$

$$\mathbf{h} \leftarrow \frac{\mathbf{h}}{\sqrt{\|\mathbf{h}\|^2 + \epsilon}}$$

Выделение пешеходов

- Пройтись с некоторым шагом по исходному изображению окошком 128 x 64
- Для каждого положения окошка найти HOG дескриптор
- С помощью обученного заранее классификатора определить, есть ли в окошке человек или его часть
- Уменьшить изображение в 2 раза и повторить



SIFT

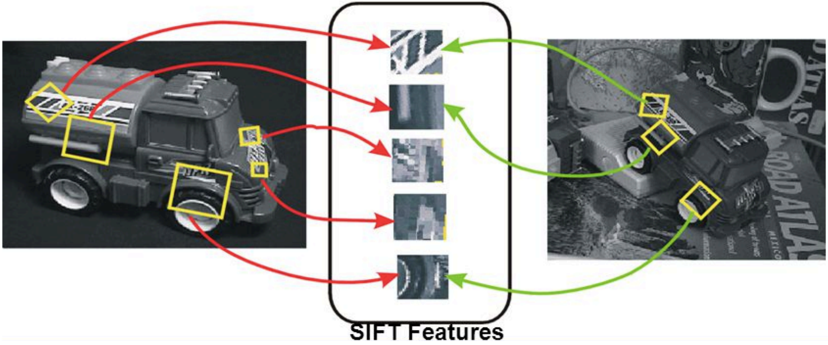
SIFT

Цель в следующем: выделить набор точек на изображении, и каждой сопоставить дескриптор, при этом дескрипторы должны быть частично или полностью инвариантны к различным преобразованиям:

SIFT

Цель в следующем: выделить набор точек на изображении, и каждой сопоставить дескриптор, при этом дескрипторы должны быть частично или полностью инвариантны к различным преобразованиям:

- Смещение
- Поворот
- Масштаб
- Зашумление
- Изменение яркости
- Изменения положения камеры


SIFT Features

Размытие по Гауссу

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2}$$

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

где $*$ - операция свертки

Размытие по Гауссу

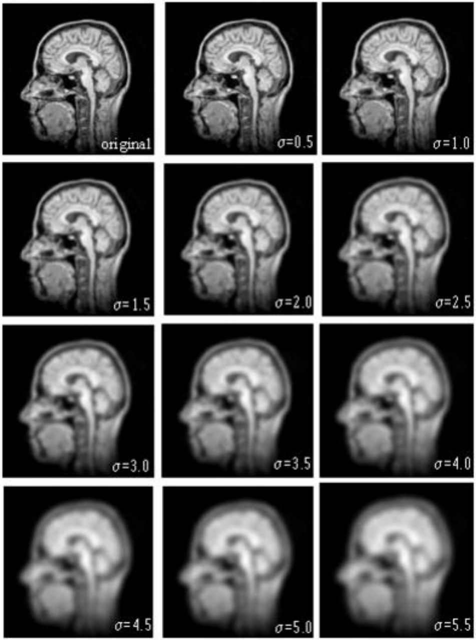
$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2}$$

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

где $*$ - операция свертки

По сути вместо значения в пикселе мы получаем взвешенное среднее в области вокруг него

$$L(x, y, \sigma) = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} G(i, j, \sigma) \cdot I(x+i, y+j)$$


original
σ=0.5
σ=1.0
σ=1.5
σ=2.0
σ=2.5
σ=3.0
σ=3.5
σ=4.0
σ=4.5
σ=5.0
σ=5.5

Пирамида разностей

- Хотим посмотреть на исходное изображение с разной степенью размытия

Пирамида разностей

- Хотим посмотреть на исходное изображение с разной степенью размытия
- Возьмем исходное σ и $k = 2^{1/s}$, построим $G(x, y, \sigma)$, $G(x, y, k\sigma)$, $G(x, y, k^2\sigma)$, ..., $G(x, y, 2\sigma)$

Пирамида разностей

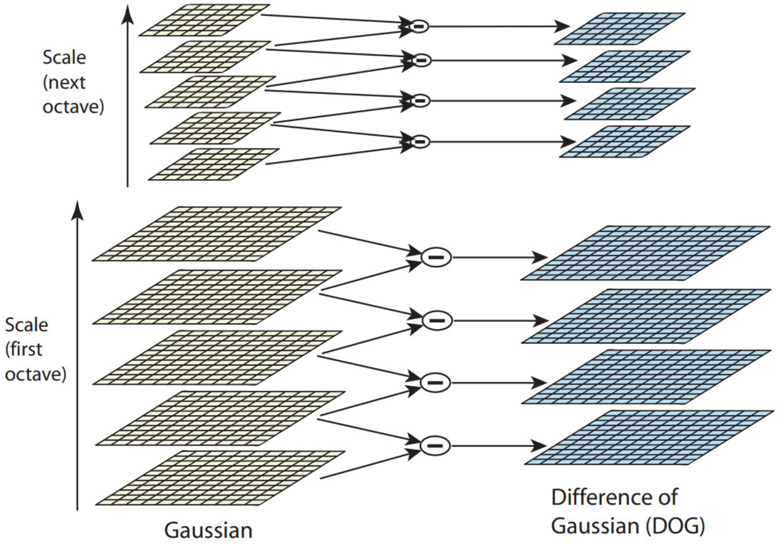
- Хотим посмотреть на исходное изображение с разной степенью размытия
- Возьмем исходное σ и $k = 2^{1/s}$, построим $G(x, y, \sigma)$, $G(x, y, k\sigma)$, $G(x, y, k^2\sigma)$, ..., $G(x, y, 2\sigma)$
- Перейдем к новой "октаве": уменьшим последнее изображение в 2 раза (выкинем каждую 2 строку и стобец), и повторим предыдущий шаг

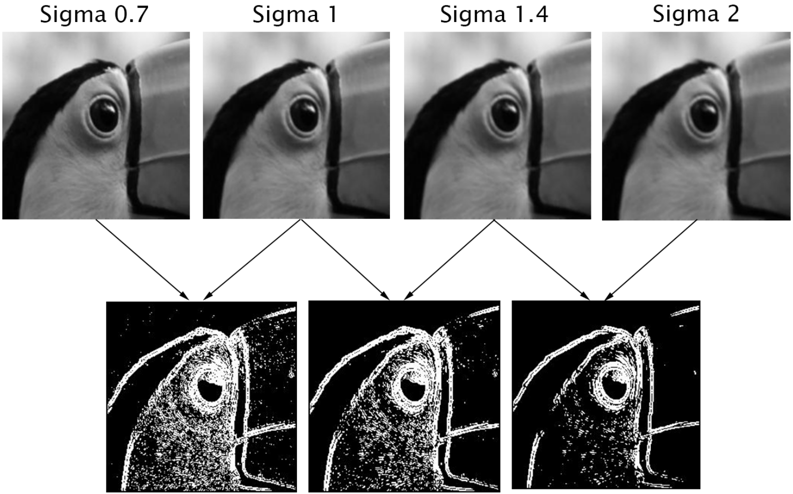
Пирамида разностей

- Хотим посмотреть на исходное изображение с разной степенью размытия
- Возьмем исходное σ и $k = 2^{1/s}$, построим $G(x, y, \sigma)$, $G(x, y, k\sigma)$, $G(x, y, k^2\sigma)$, ..., $G(x, y, 2\sigma)$
- Перейдем к новой "октаве": уменьшим последнее изображение в 2 раза (выкинем каждую 2 строку и стобец), и повторим предыдущий шаг
- В каждой октаве построим разности соседей

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) =$$

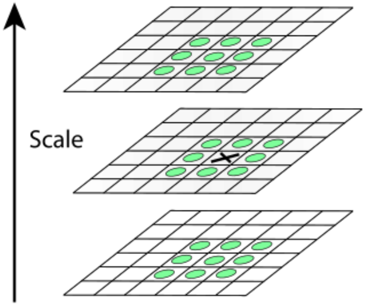
$$= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$$


Scale
(next
octave)
Scale
(first
octave)
Gaussian
Difference of
Gaussian (DOG)



Difference of Gaussian

Экстремумы



- Кандидаты в ключевые точки - локальные экстремумы в пирамиде разностей
- С каждой точкой-кандидатом связан ее масштаб в пирамиде, тем самым достигается инвариантность относительно масштаба

Локализация

Определим положение ключевой точки с субпиксельной точностью:

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

Где $x = (x, y, \sigma)^T$. Для определения точки экстремума приравняем градиент функции к нулю, получим

Локализация

Определим положение ключевой точки с субпиксельной точностью:

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

Где $x = (x, y, \sigma)^T$. Для определения точки экстремума приравняем градиент функции к нулю, получим

$$\frac{\partial D^T}{\partial x} + \frac{1}{2}\left(\frac{\partial^2 D}{\partial x^2} + \left(\frac{\partial^2 D}{\partial x^2}\right)^T\right)x = 0$$

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

Локализация

Определим положение ключевой точки с субпиксельной точностью:

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

Где $x = (x, y, \sigma)^T$. Для определения точки экстремума приравняем градиент функции к нулю, получим

$$\frac{\partial D^T}{\partial x} + \frac{1}{2}\left(\frac{\partial^2 D}{\partial x^2} + \left(\frac{\partial^2 D}{\partial x^2}\right)^T\right) x = 0$$

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

Если сдвиг $\hat{x}$ больше 0.5 по какой-то координате, надо сдвинуть исходную точку в этом направлении.
Также отбросим точки с низким контрастом, то есть точки, где $|D(\hat{x})| < t$.

Исключение ребер

Посчитаем матрицу Гессе в ключевой точке

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

Исключение ребер

Посчитаем матрицу Гессе в ключевой точке

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

Пусть α - большее по модулю собственное значение, β - меньшее, также $\alpha = r\beta$

Исключение ребер

Посчитаем матрицу Гессе в ключевой точке

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

Пусть α - большее по модулю собственное значение, β - меньшее, также $\alpha = r\beta$

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

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

Исключение ребер

Посчитаем матрицу Гессе в ключевой точке

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

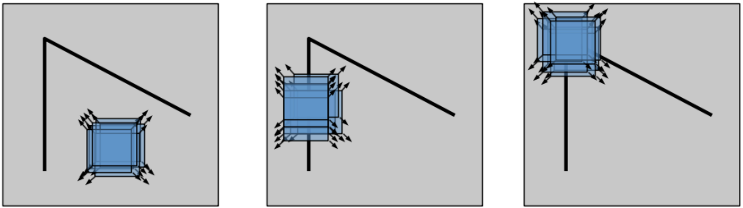
Пусть α - большее по модулю собственное значение, β - меньшее, также $\alpha = r\beta$

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

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

Тогда мы фиксируем некоторое r_0 (например 10), и говорим, что нас интересует только те точки, где

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r_0+1)^2}{r_0}$$



“flat” region:
no change in all
directions

“edge”:
no change along the
edge direction

“corner”:
significant change in
all directions

Ориентация

Для достижения инвариантности отностиельно вращения присвоим каждой ключевой точке направление:

Ориентация

Для достижения инвариантности отностиельно вращения присвоим каждой ключевой точке направление:

- В соответствующем точке Гауссиане с масштабом σ построить гистограмму направленных градиентов на 36 ячейках (покрывая 360 градусов), при этом вклад каждого вектора домножается на функцию Гаусса с масштабом 1.5σ в соответствующей точке.

Ориентация

Для достижения инвариантности отностиельно вращения присвоим каждой ключевой точке направление:

- В соответствующем точке Гауссиане с масштабом σ построить гистограмму направленных градиентов на 36 ячейках (покрывая 360 градусов), при этом вклад каждого вектора домножается на функцию Гаусса с масштабом 1.5σ в соответствующей точке.
- Точке присваивается направление соответствующее ячейке с максимальным весом

Ориентация

Для достижения инвариантности отностиельно вращения присвоим каждой ключевой точке направление:

- В соответствующем точке Гауссиане с масштабом σ построить гистограмму направленных градиентов на 36 ячейках (покрывая 360 градусов), при этом вклад каждого вектора домножается на функцию Гаусса с масштабом 1.5σ в соответствующей точке.
- Точке присваивается направление соответствующее ячейке с максимальным весом
- Если есть ячейки с весом хотя бы 0.8 от максимального, создаются дубликаты ключевой точки с соответствующими направлениями

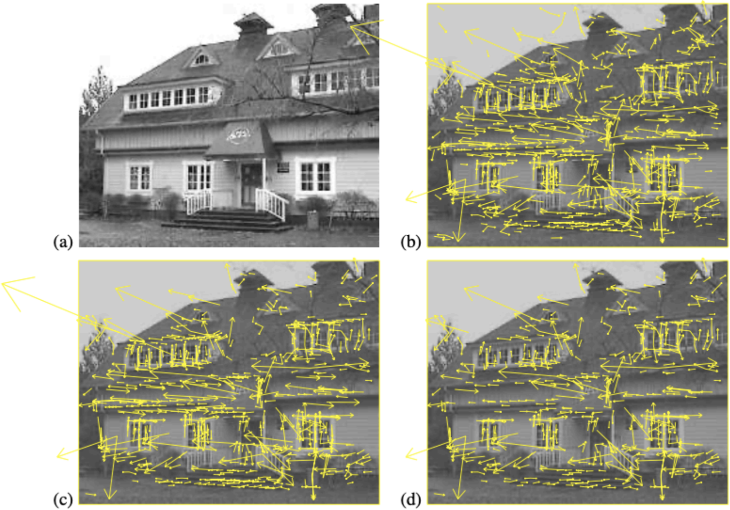
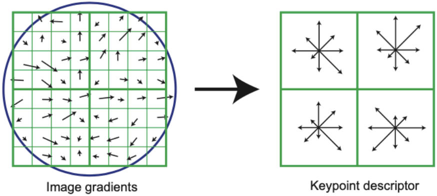


Figure 5: This figure shows the stages of keypoint selection. (a) The 233x189 pixel original image. (b) The initial 832 keypoints locations at maxima and minima of the difference-of-Gaussian function. Keypoints are displayed as vectors indicating scale, orientation, and location. (c) After applying a threshold on minimum contrast, 729 keypoints remain. (d) The final 536 keypoints that remain following an additional threshold on ratio of principal curvatures.

Дескриптор

- Взять (квадратную) область вокруг ключевой точки в соответствующем Гауссиане масштаба σ
- Повернуть ее и градиенты всех клеток внутри на угол, соответствующий направлению ключевой точки
- Разбить область на 4 блока, в каждом построить гистограмму направленных градиентов (взвешенных функцией Гаусса)
- Сложить гистограммы в итоговый вектор и нормировать его (как в HOG)



Источники

HOG:

- https://www2.cs.duke.edu/courses/fall15/compsci527/notes/hog.pdf
- https://www.learnopencv.com/histogram-of-oriented-gradients/
- https://www.youtube.com/watch?v=0Zib1YEE4LU
- https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients

SIFT:

- https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf
- https://habr.com/ru/post/106302/
- https://www.youtube.com/watch?v=NPcMS49V5hg
- https://en.wikipedia.org/wiki/Scale-invariant_feature_transform

Спасибо за внимание!