SIFT(SCALE INVARIANCE FEATURE TRANSFORM)

Введение

SIFT - Масштабно-инвариантный дескрипторв ключевых областей на изображении. Такой десриптор используется в задачах связанных с оптическими потоками, такими как:

- Реконструкция аффиных преобразований (вращающих и масштабирующих пространство)
- Создание панорам
- Реконструкция трехмерного объекта по его двумерным проекциям
- Слежение за движением объекта по снимкам

Целью SIFT является извлечение ключевых точек на изображениях. На английском такие точки называют “Interesting points”. Другим вопросом являются критерии, которым должна отвечать ключевая точка. Свойства такой точки следующие:

1. Окрестность такой точки высоко информативна(Велика дисперсия интенсивности, цветовая дисперсия)
2. Достаточно репрезентативна, чтобы сопоставить с другими точками
3. Имеет четко определяемую позицию (Отличимость от других точек)
4. Масштабно и ротационно инварианта
5. Нечувствительной к параметрам освещения

Чтобы извлечь такие точки, нам необходимо некоторое преобразование, такое преобразование должно отвечать следующим критериям:

- Инвариантно относительно смещения
- Инвариантно относительно поворота
- Инвариантно относительно масштаба (точка должна быть ключевой и с заданным индексом на двух изображениях с разным масштабом)
- Изменение яркости

Gaussian scale space

Можно ли найти такое ядро G_σ, которое с различным параметром σ будет представлятт изображение $f(x,y)$ в пространсте масштаба $L(x,y,\sigma)$ путем сворачивания G_σ и $f(x,y)$, то есть $L(x,y,\sigma) = (G_\sigma * f)(x,y)$ и будет удовлетворять следующим аксиомам пространста масштаба?

1. Линейность: $G_\sigma * (af + bh) = aG_\sigma * f + bG_\sigma * h$
2. Инвариантность к сдвигу: $G_\sigma * S_{(\triangle x, \triangle y)} * f = S_{(\triangle x, \triangle y)} * G_\sigma * f$ (Порядок действия операторов не важен. либо бы мы сместим изображение на $(\triangle x, \triangle y)$ и подействуем ядром G_σ, либо мы подействуем ядром G_σ и сместим изображение на $(\triangle x, \triangle y)$ получим одно и тоже)
3. Структура полугруппы или каскадное сглаживание. $g(x,y,\sigma_1) * g(x,y,\sigma_2) = g(x,y,\sigma_1 + \sigma_2)$
4. Гладкость: $\partial_\sigma L(x,y,\sigma) = (D * L)(x,y,\sigma)$. Оператор G_σ не должен формировать новых максимумов или множить их
5. Инвариантность к вращению: $g(x,y,\sigma) = h(x^2 + y^2, \sigma)$ для любой функции h(изотропность)
6. Масштабная инвариантность $F(u,v,\sigma) = F\left(\frac{u}{\varphi(\sigma)}, \frac{v}{\varphi(\sigma)}\right)$, где F преобразование Фурье
7. Свойства нормы $\iint_{x\in(-\infty,\infty), y\in(-\infty,\infty)} G(x,y,\sigma)dxdy = 1$

Было доказано, что такому пространству соответствует Гауссово ядро, причем принято считать, что константа σ в знаменателе коэффициента усиления опущена:

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

Интуиция заключается в том, что, Гауссов фильтр - фильтр ФНЧ. Отсюда следует, что он удаляет высокочастотные детали изображения. Очевидно, что чем мы дальше от изображения, тем меньше деталей можно увидеть. Таким образом, сглаживая изображение, мы можем думать о результате сглаживания как о версии исходного изображения в отдалении. Особенно это будет ясно тогда, когда после сглаживания мы проведем децимацию, изображение станет меньшей размерности, число пикселей уменьшиться и оно само по себе станет меньше. Тем не менее связь между изображениями с разными σ видима и ощутима:

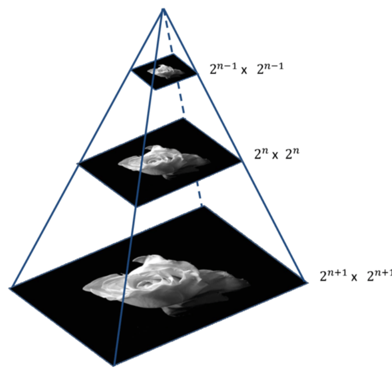


Figure 1: Пирамида Гауссова пространства масштаба

Лапласиан Гауссиана и детекция ключевой точки

По сути вещей поиск ключевой точки а алгоритме SIFT является поиском экстремума внутри контурного участка.

Первый вариант:

Первая производная Гауссиана дает такой оператор, который не отвечает, как минимум, свойству инвариантности к вращению, с другой стороны LoG или так называемый вейвлет - мексиканская шляпа отвечает всем аксиомам пространства масштаба. Поэтому такое ядро используется для детектирования ключевых точек.

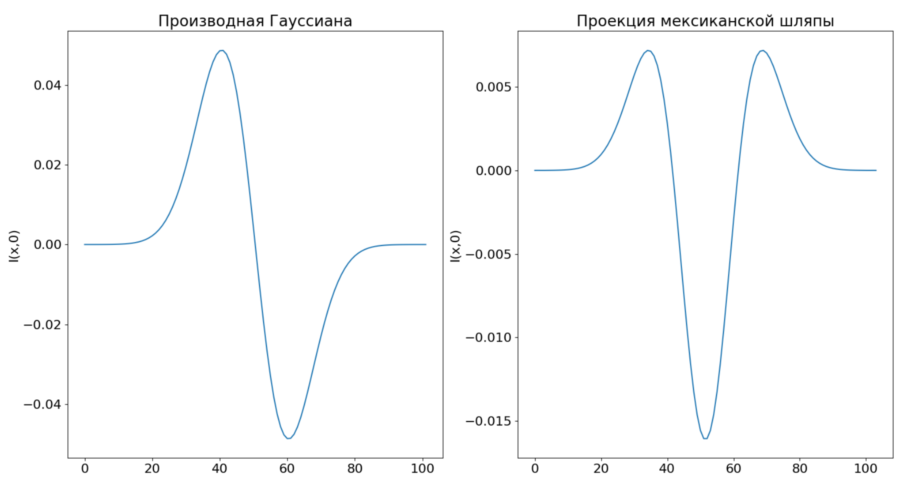


Figure 2: Проекция дифференциала Гауссиана и инвариантная к повороту проекция мексианская шляпа

$$\Delta G(x, y, \sigma) = \frac{\partial^2 G(x, y, \sigma)}{\partial x^2} + \frac{\partial^2 G(x, y, \sigma)}{\partial y^2} = \frac{1}{\pi\sigma^2} \left(\frac{x^2 + y^2}{2\sigma^2} - 1 \right) exp \left(-\frac{x^2 + y^2}{2\sigma^2} \right) \tag{2}$$

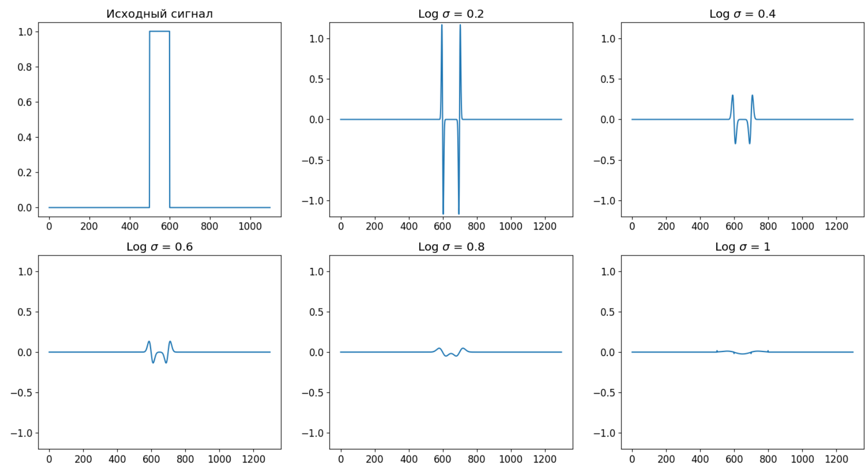


Figure 3: Свертка прямоугольного импульса с LoG оператором при различных σ

Таким образом, отсюда следует, что невозможно определить ключевую точку и масштаб σ,при подействовании на сигнал оператором (2). Однако, если нормировать Лапласиан Гауссиана, так чтобы пространство было также мастабно инвариантным, мы можем найти σ и определить максимумы, в которых располагаются ключевые точки:

$$\frac{1}{\pi\sigma^2}\left(\frac{x^2+y^2}{2\sigma^2}-1\right)exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)=\frac{k}{\pi s^2\sigma^2}\left(\frac{s^2x^2+s^2y^2}{2s^2\sigma^2}-1\right)exp\left(-\frac{s^2x^2+s^2y^2}{2s^2\sigma^2}\right)$$

Мы получаем одно и тоже в случае, если $k=s^2$. Таким образом нормированный 2D LoG имеет вид:

$$\Delta_{norm}G(x,y,\sigma)=s^2\Delta_{norm}G(sx,sy,s\sigma) \tag{3}$$

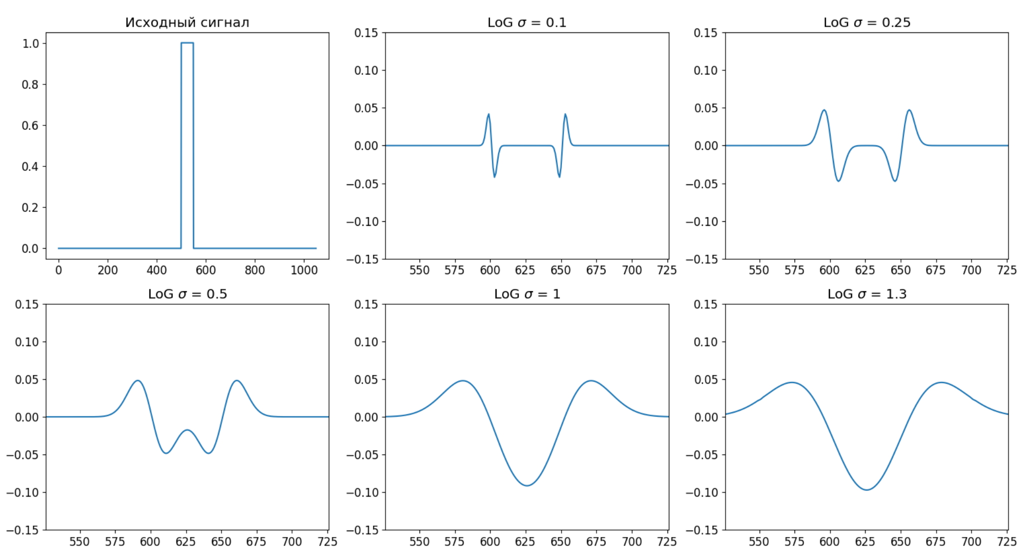
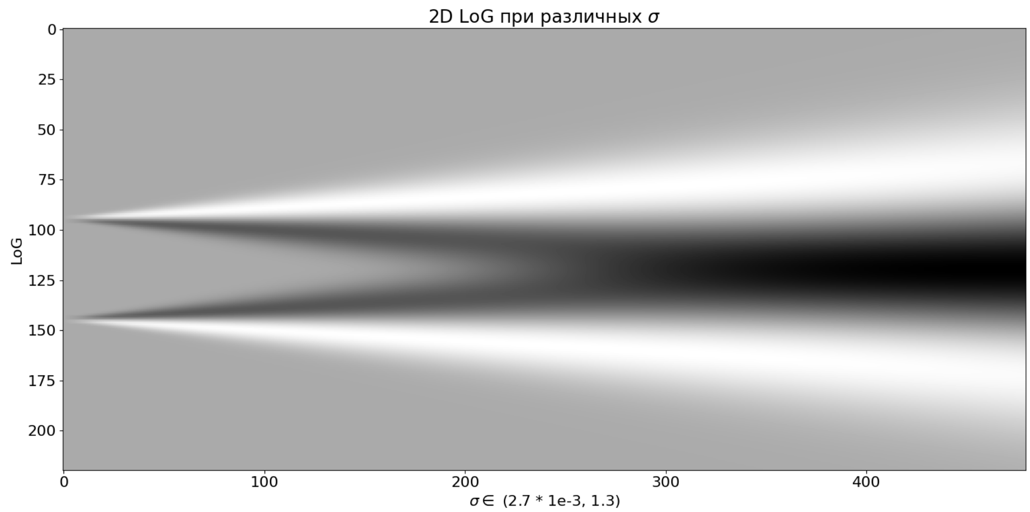


Figure 4: Свертка прямоугольного импульса мастабно инвариантного LoG при различных масштабах σ

Отсюда следует, что используя (3) как ядро для импульсной детекции (blob detection problem), мы в результате можем получить как значение масштаба σ так и координату центра как разность между максимумом результата свертки LoG и изображения и половиной длины окна LoG.



.

Figure 5: 2D LoG

Поиск LoG является вычислительно сложным. Другим, вычислительно более экономным способом является применение разности Гауссианов, известное как DoG. DoG - аппроксимация LoG на основе уравнения теплопроводности.

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2 \Delta G(x, y, \sigma) \tag{4}$$

Для расчета (4) мы используем концепцию пирамид.

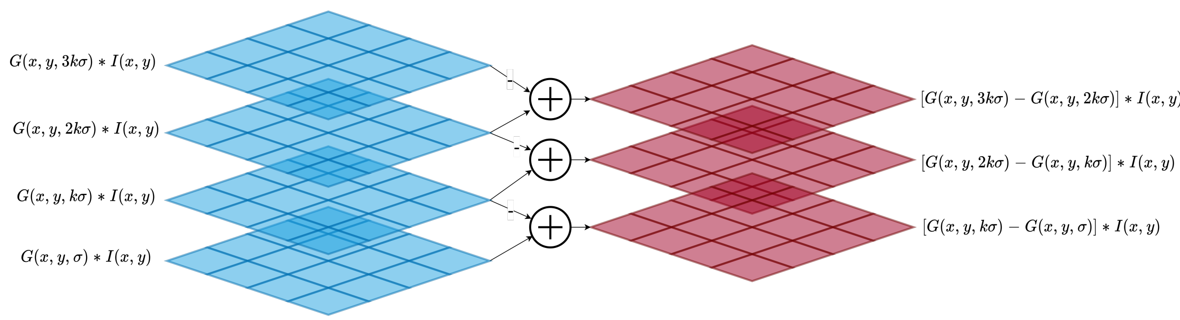


Figure 6: Графическая интерпретация стеков DoG, которые в результате формируют Лапласиан Гауссиана для одного элемента из пирамиды

Для изображения свернутого с $G(x, y, k\sigma)$ для 4 разных σ получаются 3 изображения, в которых ищется локальный максимум в покоординатно одинаковых окнах среди всех соседей. Принято использовать окно размерности $3x3$, что потребует сравнивать каждую точку с 8-ю соседями на заданном уровне, 9 соседями сверху и 9 соседями снику: Таким образом каждую точку в окне необходимо сравнить с $8 + 9 + 9 = 26$ соседями. Если такая точка является экстремумом, то она является потенциальной ключевой точкой.

Валидация и исключение ключевых точек

Поскольку работа ведется на дискретном множестве значений и полученные точки определены всего для 4 разных масштабов σ в пределах одной октавы, результат DoG фильтрации требует более точной локализации экстремумов. Мы можем разложить функцию DoG в ряд Тейлора, получить новую функцию и пересчитать экстремумы. Известно, что для векторов ряд Тейлора имеет вид:

$$T(x) = \sum_{\alpha} \frac{(x-a)^{\alpha}}{\alpha!} \left(\partial^{\alpha} f\right)(a) \tag{5}$$

В нашем случае, достаточно использовать разложение DoG в ряд Тейлора 2 порядка в точке экстремума:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} \frac{\partial^2 D}{\partial x^2} x \tag{6}$$

Раскладывая DoG в ряд Тейлора 2 порядка в окрестности точки экстремума $x = (x, y, \sigma)$ (или a в (5)), приравнивая к нулю мы можем определить уточненный экстремум $\hat{x} = (\hat{x}, \hat{y}, \hat{\sigma})$ на множестве $\mathbb{R}$:

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \tag{7}$$

Если одна из компонент вектора $\hat{x} = (\hat{x}, \hat{y}, \hat{\sigma})$ больше чем 0,5*шаг_сетки_в_этом_направлении, то мы смещаемся к новой точке в этом направлении. Для новой соседней точки выполняется все то же самое. Если мы таким образом вышли за пределы октавы, то эту точку исключают из рассмотрения.

Следующая проверка заключается в исключении точек, которые находятся на границе какого-то объекта или плохо освещена, то такую точку исключают из рассмотрения. По сути вещей исключают точки с сильной ассиметрией, то есть те точки, которые мало похожи на изотропные импульсы, поскольку высококоррелированные края могут появиться в двух разных изображениях, но при этом на изображении будет видно, что это совсем разные области.

Изгибы можно определяются Гессианом, в который включена не только степень кривизны, но и ее направление, что по существу нас и интересует:

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

Решая задачу относительно собственных значений можно увидеть, что λ_{max}соответствует направлению максимальной кривизны а λ_{min}соответствует направлению минимальной кривизны. Нашей же целью является избавиться от информации о величинах кривизны, но оставить их отношение. Известно, что $TrH = \lambda_{max} + \lambda_{min}$, а $detH = \lambda_{max} * \lambda_{min}$. Выразив $\lambda_{max} = r\lambda_{min}$и найдя квадрат отношения следа к определителю, мы можем избавиться от собственных значений, оставив только отношение r

$$\frac{(TrH)^2}{detH} = \frac{(r+1)^2}{r} \tag{8}$$

Таким образом сравнивая $\frac{(TrH)^2}{detH}$ с $\frac{(r+1)^2}{r}$ мы можем понять, лежит ли область с экстремумом в классе с импульсом или просто является краем. Обычно принято считать, что оптимальное значение $\frac{(r+1)^2}{r} = 10$

Нахождение ориентации ключевой точки

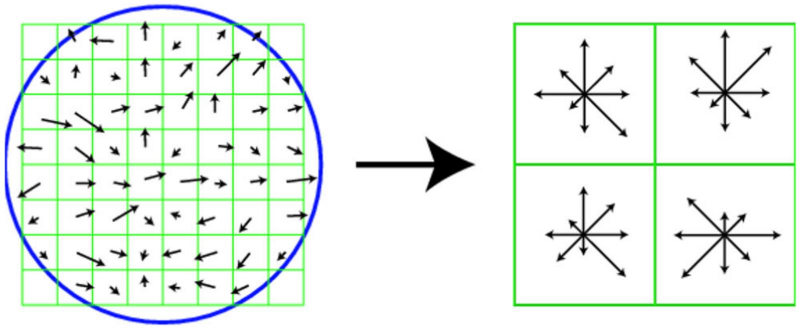
Качество сопоставлении без информации о направлении ключевой точки значительно снижается. Принято использовать статистический метод для оценки направления ключевой точки.

Algorithm 1 Оценка направления ключевой точки

- В окне с центром в ключевой точке, рассчитать величины и направления градиентов используя DoG
- Расчитать амплитуду и фазу градиентов
- Пороговым методом исключить из рассмотрения фазы точек в которых модуль градиента мал.
- Построить гистограмму фаз
- Считать ориентацией ключевой точки центр бина с максимальным значением

Дескрипторы ключевых точек

Когда ключевые точки и их направления определены, для сопоставления изображений, необходимо определить дискрипторы или уникальные идентификаторы точек. По сути вещей идентификатор это просто число (Оно без контекста не имеет смысла), которое сопоставимо к объекту, но дескриптор также включает какую-то дополнительную информацию. Для запатентованного SIFT детектора дескриптор имеет размерность: 4x4x8=128 точек.



(a) image gradients (b) keypoint descriptor

Figure 7: Исходная область и визуальная интерпретация дескриптора 2x2x8

Перед расчетами окно, в котором рассчитываются дескрипторы, поворачивают в направлении градиента. В результате это пользволяет достичь ротационной инвариантности.

4x4 - обозначают размерность дескриптора, а 8 говорит о размерности гистограммы. Такая гистограмма строится из всех векторных направлений в подобласти дескриптора. Поскольку направление, ближайшее к ключевой точке имеет наибольше значение, подобласть поточечно взвешивают Гауссовым окном с σ равной половине окна дескриптора. Результирущий вектор нормализуется, и все значения меньше 0,2 приравниваются к 0,2. Затем гистограмма повторно нормализуется, что позволяет достичь устойчивости относительно к аффинным изменениям овещения.

Algorithm 2 Поиск дескриптора ключевой точки

1. Извлечь окно в окрестности ключевой точки.
2. Повернуть окно в направлении ориентации ключевой точки
3. Разделить окно на подобласти и для каждой:
4. Взвесить подобласть Гауссовым окном
5. Получить гистограммы направлений
6. Нормализовать гистограмму к 1
7. Приравнять все значения меньшие 0,2 к 0,2
8. Снова нормализовать гистограмму к 1

Вывод

Недостатком SIFT является то, что не все полученные дескрипторы будут отвечать требованиям. Не на всех изображениях SIFT будет квалифицированно находить ключевые точки. Например, при поиске аффиных преобразований двух изображений кирпичной стены, метод работать не будет, так как он оперирует с уникальностью объектов.

Был рассмотрен SIFT - дескриптор - масштабно-инвариантный дескриптор ключевых точек. Области детекции, которые отвечают требованиям дескриптора называют blobs - сгустки или импульсы. Центры таких импульсов можно найти с помощью нормированного Лапласиана Гауссиана, который в $\mathbb{R}$ дает как значение масштаба σ, так и координату экстремума на изображении. Такая точка экстремума является не только величиной масштаба σ, но и координатой центра импульса на изображении. Поскольку изображение лежит в дискретно-квантованном пространстве, более экономным методом вычисления вектора (x, y, σ) является применение Разности Гауссианов, боеле известных как DoG, для дискретного множества σ, состоящего из 4 значений. В результате это дает нам три дважды продифференцированных изображения для 4 масштабов. На самом деле такого множества недостаточно для точного определения σ, поэтому используется концепция пирамид, в которых получается несколько таких групп по 3 изображения. Причем с ростом мастштаба σ, для экономии вычислительных ресурсов следующая тройка изображений децимируется в два раза. Накладывая децимированные изображения от меньшей размерности к большему, получаем стек, который визуально напоминает пирамиду.

Сам поиск ключевой точки определяется его сравнением с соседями. Если его 8 соседей на уровне и по 9 на верхнем и по 9 на большем меньше него, то такая точка считается кандидатом на экстремальность. Далее осуществляется проверка кандидатов на правильность найденной точки а также на соответствие формы.

Кандидатов проверяют разложением в ряд Тейлора 2 порядка в окрестности с их координатой. Используя аппроксимацию мы получаем новую функцию и ищем экстремум на ней. Если экстремальная точка находится ближе к другой точки пирамиды, то другая считается новым экстремумом. И также проверяется она, следующая и т.д. до тех пор пока ключевая точка не будет найдена.

Вторая проверка, проверка на форму импульса основана на Гессиане. Известно, что собственные значения Гессиана второго порядка для объекта дифференциальной геометрии представляют собой значения с максимальной и минимальной кривизны. Если отношение квадрата следа Гессиана и определителя Гессиана меньше какого то значения, то такая точка больше похожа на импульс и не является краем. Таким образом все кандидаты большие по значению заданного отношения исключаются.

Повысить уникальность экстремальной точки можно включив в нее информацию о направлении градиента. Для этой задачи используется статистический градиент, который берет окрестность экстремума, и формирует гистограмму углов с направлением градиента, для тех θ, в которых модуль градиента существенен или больше некоторого значения. Результирующим направлением является тот угол θ, значение которого максимально для рассчитаной гистограммы

После того, как найдены ключевые точки, для них формируются дескрипторы, которые являются "уникальными описаниями", что позволяют сопоставить две точки на разных изображениях. Дескриптор расчитывается из точек изображения 16x16, которые делятся на группы 4x4. В каждой группе есть по 8 значений углов градиента представленных в виде гистограмм. Каждая гистограмма нормализуется, в ней приравниваются к 0.2 значения меньшие 0.2, затем гистограмма нормализуется снова.