SIFT

the Scale Invariant Feature Transform

Distinctive Image Features from Scale-Invariant Keypoints (springer.com)

cascade filtering 级联滤波降低运算开销

根据Euclidean distance of their feature vectors 匹配

使用fast nearest-neighbor algorithms （快速最近邻）

a least-squared estimate is made for an affine approximation to the object pose 最小二乘拟合仿射变换

生成图像特征集合主要步骤：

1. Scale-space extrema detection 尺度空间极值检测

   difference-of-Gaussian function(DOG)

2. Keypoint localization 关键点定位
3. Orientation assignment 方向配赋
4. Keypoint descriptor 关键点描述

Detection of Scale-Space Extrema

a continuous function of scale known as scale space

the only possible scale-space kernel is the Gaussian function

尺度空间定义：

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

the convolution of a variable-scale Gaussian, G(x, y, σ), with an input image, I (x, y).

差分：

the difference-of-Gaussian function convolved with the image, D(x, y, σ):

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

the difference-of-Gaussian function provides a close approximation to the scale-normalized Laplacian of Gaussian, $\sigma^2\nabla^2 G$

from the heat diffusion equation:

$$\sigma\nabla^2 G = \partial G/\partial\sigma \approx (G(x, y, k\sigma) - G(x, y, \sigma))/(k\sigma - \sigma)$$

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2\nabla^2 G.$$

when the difference-of-Gaussian function has scales differing by a constant factor it already incorporates the σ^2 scale normalization required for the scale-invariant Laplacian.

The factor (k − 1) in the equation is a constant over all scales and therefore does not influence extrema location.

Local Extrema Detection

局部极值检测

In order to detect the local maxima and minima of D(x, y, σ), each sample point is compared to its eight neighbors in the current image and nine neighbors in the scale above and below

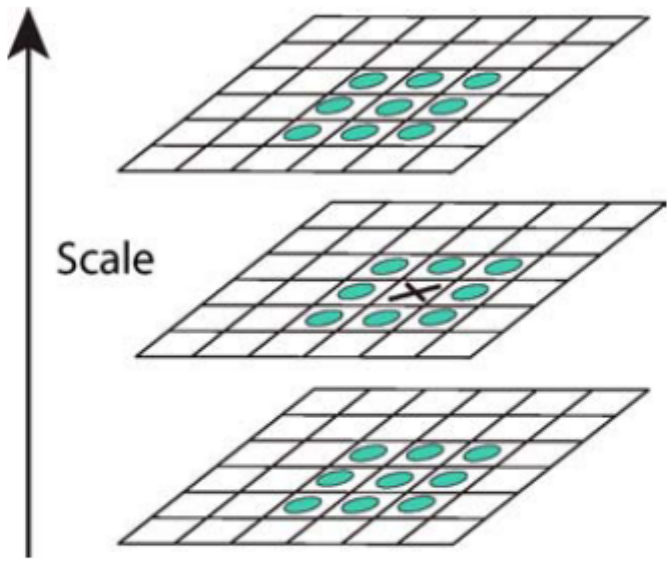

Figure 2. Maxima and minima of the difference-of-Gaussian images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3 × 3 regions at the current and adjacent scales (marked with circles).

Frequency of Sampling in Scale

抽样频率

the scale-space difference-of-Gaussian function has a large number of extrema and that it would be very expensive to detect them all.

Frequency of Sampling in the Spatial Domain

空间域采样频率

the repeatability continues to increase with σ .

there is a cost to using a large σ in terms of efficiency.

we have chosen to use σ = 1.6, which provides close to optimal repeatability

to make full use of the input, the image can be expanded to create more sample points than were present in the original.

We double the size of the input image using linear interpolation prior to building the first level of the pyramid.在建立金字塔第一层前，线性插值使输入规模翻倍。

Accurate Keypoint Localization

Once a keypoint candidate has been found by comparing a pixel to its neighbors, the next step is to perform a detailed fit to the nearby data for location, scale,and ratio of principal curvatures.

比较像素点与其相邻像素得到关键点候补后，对其附近像素进行拟合。

fitting a 3D quadratic function to the local sample points to determine the interpolated location of the maximum

三维二次函数拟合局部样本，确定插值最大值位置。

the Taylor expansion (up to the quadratic terms) of the scale-space function, D(x, y, σ)

D(x, y, σ)泰勒展开(二次)：

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

x = (x, y, σ)T is the offset from this point

求导得到极值点：

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

the Hessian and derivative of D are approximated by using differences of neighboring sample points. The resulting 3×3 linear system can be solved with minimal cost.

相邻样本点差分逼近

Eliminating Edge Response

消除边际效应

The difference-of-Gaussian function will have a strong response along edges, even if the location along the edge is poorly determined and therefore unstable to small amounts of noise.

A poorly defined peak in the difference-of-Gaussian function will have a large principal curvature across the edge but a small one in the perpendicular direction.

The principal curvatures can be computed from a 2 × 2 Hessian matrix, H

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

取相邻样本点差值估计导数

The eigenvalues of H are proportional to the principal curvatures of D. H特征值

Let α be the eigenvalue with the largest magnitude and β be the smaller one.

α幅值最大特征值，β幅值较小特征值

$$Tr(H) = Dxx + Dyy = \alpha + \beta, Det(H) = DxxDyy - (Dxy)^2 = \alpha\beta.$$

$$\alpha = r\beta$$

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

检查主曲率之比是否小于阈值r,只需检查：

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

剔除主曲率之比过大的关键点。

Orientation Assignment

By assigning a consistent orientation to each keypoint based on local image properties, the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.

通过局部属性为每个关键点分配一致的方向，关键点描述基于该方向。

The scale of the keypoint is used to select the Gaussian smoothed image, L, with the closest scale, so that all computations are performed in a scale-invariant manner. For each image sample, L(x, y), at this scale, the gradient magnitude, m(x, y), and orientation, θ(x, y),is precomputed using pixel differences:

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

An orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint.

直方图峰值对应局部梯度优势方向。

对于具有多个相近峰值的位置，在相同位置和尺度上创建多个不同方向关键点。

Only about 15% of points are assigned multiple orientations, but these contribute significantly to the stability of matching.

The Local Image Descriptor

The next step is to compute a descriptor for the local image region that is highly distinctive yet is as invariant as possible to remaining variations, such as change in illumination or 3D viewpoint.

Descriptor Representation

First the image gradient magnitudes and orientations are sampled around the keypoint location,using the scale of the keypoint to select the level of Gaussian blur for the image.

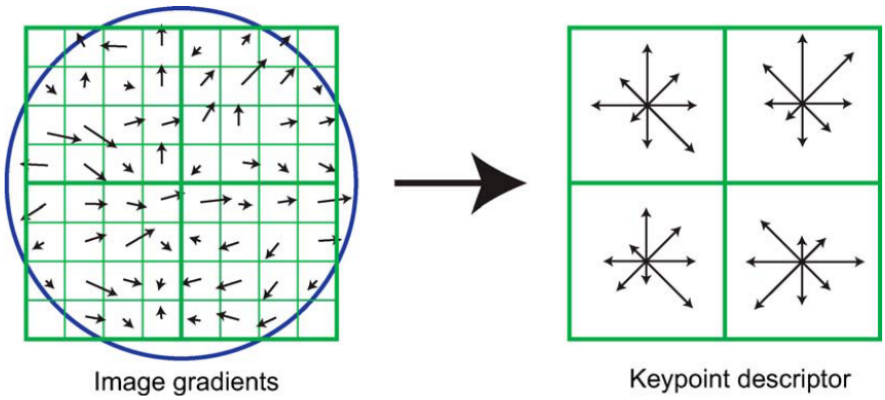

Figure 7. A keypoint descriptor is created by first computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location, as shown on the left. These are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then accumulated into orientation histograms summarizing the contents over 4x4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. This figure shows a 2×2 descriptor array computed from an 8×8 set of samples, whereas the experiments in this paper use 4×4 descriptors computed from a 16×16 sample array.

A Gaussian weighting function with σ equal to one half the width of the descriptor window is used to assign a weight to the magnitude of each sample point.

trilinear interpolation is used to distribute the value of each gradient sample into adjacent histogram bins. 三线性插值

each entry into a bin is multiplied by a weight of 1−d for each dimension, where d is the distance of the sample from the central value of the bin as measured in units of the histogram bin spacing.

Descriptor Testing

the number of orientations, r 方向数

the width, n, of the n × n array of orientation histograms宽度

The size of the resulting descriptor vector is rn^2.描述符大小

Sensitivity to Affine Change

描述符对仿射变换敏感性

additional SIFT features are generated from 4 affine-transformed versions of the training image corresponding to 60 degree viewpoint changes.

Matching to Large Databases

Application to Object Recognition

Keypoint Matching

The best candidate match for each keypoint is found by identifying its nearest neighbor in the database of keypoints from training images. The nearest neighbor is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector

最近邻（描述向量最小欧式距离）

A more effective measure is obtained by comparing the distance of the closest neighbor to that of the second-closest neighbor.

The probability density functions（PDF） for correct and incorrect matches are shown in terms of the ratio of closest to second-closest neighbors of each keypoint.

匹配正确概率：最近邻与次近邻的比值

Efficient Nearest Neighbor Indexing

近似算法：the Best-Bin-First (BBF) algorithm

The BBF algorithm uses a modified search ordering for the k-d tree algorithm so that bins in feature space are searched in the order of their closest distance from the query location. 改进k-d树算法的搜索顺序

This search order requires the use of a heap-based priority queue for efficient determination of the search order. 使用基于堆的优先队列

Clustering with the Hough Transform

霍夫变换聚类

We have found that reliable recognition is possible with as few as 3 features.

只需3个特征就可实现可靠识别。

The Hough transform identifies clusters of features with a consistent interpretation by using each feature to vote for all object poses that are consistent with the feature.

Each of our keypoints specifies 4 parameters:2D location, scale, and orientation, and each matched keypoint in the database has a record of the keypoint's parameters relative to the training image in which it was found.

为关键点指定4个参数：2D位置，比例，方向。

Solution for Affine Parameters

仿射变换求解

The Hough transform is used to identify all clusters with at least 3 entries in a bin. Each such cluster is then subject to a geometric verification procedure in which a least-squares solution is performed for the best affine projection parameters relating the training image to the new image.

最小二乘拟合

仿射变换求解仅需3个匹配点

The affine transformation of a model point [x y]T to an image point [u v]T can be written as :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

where the model translation is [tx ty]T and the affine rotation, scale, and stretch are represented by the mi parameters.

We can write this linear system as Ax = b

最小二乘解：$x = [A^T A]^{-}1 A^T b,$