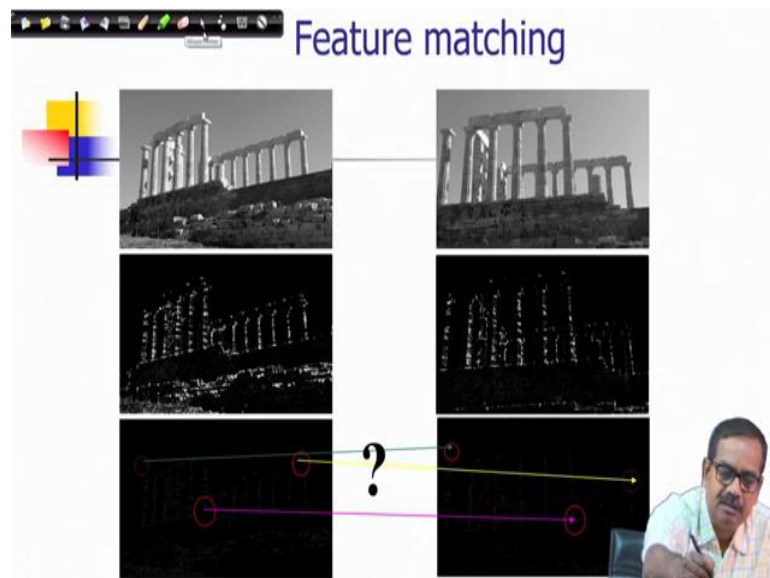


Computer Vision
Prof. Jayanta Mukhopadhyay
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 25
Feature Detection and Description Part - II

We are discussing the topic of Feature Detection and Description and in the last lecture we discussed how the interesting points or key points which will remain mostly invariant of different transformation those can be detected those are the corner points. So, the question is that now how do you characterize those corner points?

(Refer Slide Time: 00:39)



For example take these two images of the same view which I have shown in this particular lecture in the beginning of this lecture. So, here as you can see that there are many various interesting points, if I perform the same computations of Harris operator, then we can obtain the corresponding feature values and there you can see that these are the corner points which are looking little faded in this display.

But these are the corner points which are here and there are various corners. So, out of them which pairs of corner points or which pairs of key points their corresponding to the same point. So, this is the question. So, how do you get those pairs of points and how do you establish those correspondences? This is the problem of feature matching that we would be considering. So, let us consider that what are the stages are there.

(Refer Slide Time: 01:47)



Matching with Features

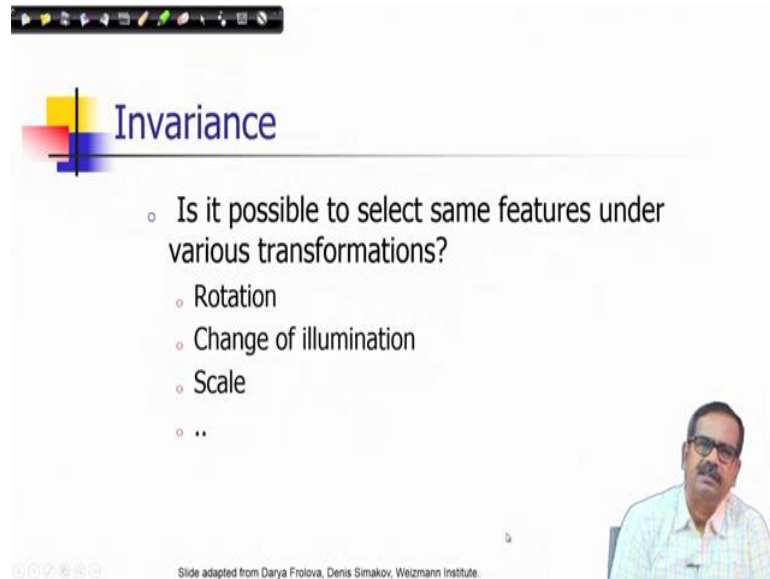
- Detect feature points in both images.
- Describe them by local statistics.
- Find corresponding pairs (Matching).

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, these are the three stages of computation of matching that we need to detect the feature points in both images. For example, we have applied the Harris operator and we considered those key points which have been extracted those are the feature points.

But we do not know that which point corresponds to which one in the other image. So, then just to uniquely describe every point we can describe them by some local statistics. So, we will discuss how the statistics could be described that is what we have to build up a descriptor for each feature point we call it feature descriptor. And then of course, using those description we have to find the corresponding pairs of points this task is known as matching.

(Refer Slide Time: 02:37)



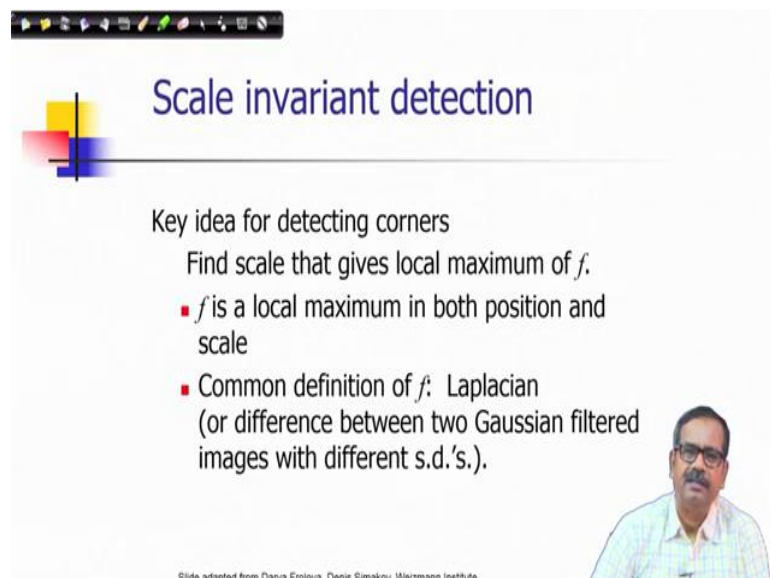
Invariance

- Is it possible to select same features under various transformations?
 - Rotation
 - Change of illumination
 - Scale
 - ..

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, again the main issue is that can you select the same features under various transformations; as we did also for consider this particular issue while detecting features this is a major concern and. Even after detecting the key points can you detect the same pair of true pairs of key points on various transformation?. So that means, your description should be also invariant to those transformation. For example, it could be rotation, change of illuminations, and variation of scale like many other such transformation.

(Refer Slide Time: 03:23)



Scale invariant detection

Key idea for detecting corners

- Find scale that gives local maximum of f .
- f is a local maximum in both position and scale
- Common definition of f : Laplacian (or difference between two Gaussian filtered images with different s.d.'s.).

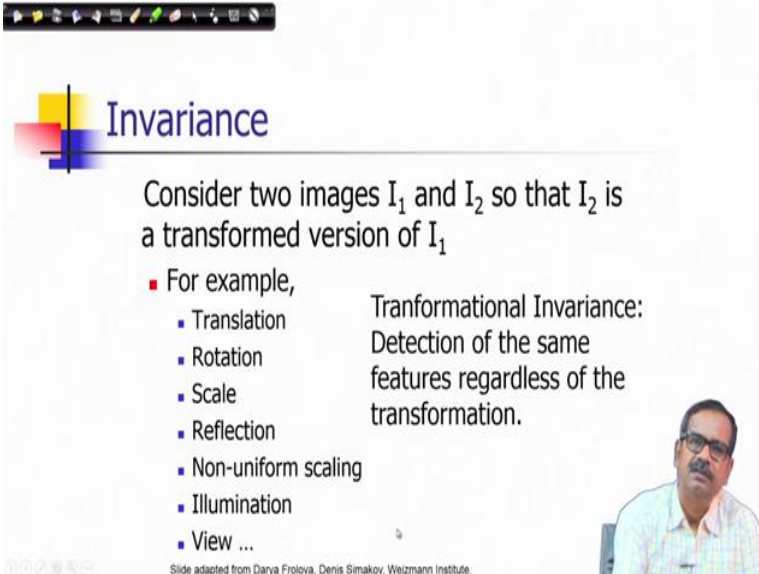
Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, the key idea for detecting the corners is that we need to find out a scale. So, one of the major concern in this detection is scale that we will find out. So, because the scale how to make the description scale independent there lies lot of tricks. Usually you can make descriptions rotation independent or translation independent that is easier to do but scale is much more trickier.

So, to get a little more scale independent what we can do or to get that scale independence features we can consider the image representation in multiple scales which means multiple resolutions. And observe that what kind of local measures of that 'f' you are considering how it varies with scales. In a proper scale it is expected that if this value which should be very high compared to the other scales which are not very appropriate for reflecting that measure. So, we will continue and we will understand this process.

So, it is a local maxima in both position and scale that is what we will be considering and there are various kinds of such measurements like Laplacian measurements which is a second derivative operations over the images. We will define this measure mathematically soon or we can consider differences between two Gaussian filtered images with different scales at different standard deviations. Standard deviations of Gaussian mask there also called scale in the image processing jargon. So, you can hear the similar terms.

(Refer Slide Time: 05:15)



Invariance

Consider two images I_1 and I_2 so that I_2 is a transformed version of I_1

- For example,
 - Translation
 - Rotation
 - Scale
 - Reflection
 - Non-uniform scaling
 - Illumination
 - View ...

Transformational Invariance:
Detection of the same features regardless of the transformation.

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, we discussed this particular thing. Let us elaborate this particular fact; say you have two images I_1 and I_2 and say I_2 is a transformed version of I_1 and as I mentioned different

kinds of transformation not only translation rotation scale it could be non-uniform scaling, it that could be illumination changes, that could be view changes, it could be reflected and so you need to get transformation invariance.

Transformational invariance of this measure which means we have to detect the same features regardless of the transformation and detection in the sense now in the description also should be unique. So, that your matching should be successful even after transformation.

(Refer Slide Time: 06:11)

Detection and description: to be invariant

Both should be ensured.

1. Detector to be transformational invariant.
 - Harris measure is invariant to translation and rotation.
 - Scale requires handling of multi-resolution representation

So, both the detection and description should be invariant and both should be ensured and as we have discussed that Harris measure is invariant to translation and rotation but we did not consider the variation over scale. So, we will see that how this could be ensured as the in the previous slide we discussed that we can use multi-resolution representation of images.

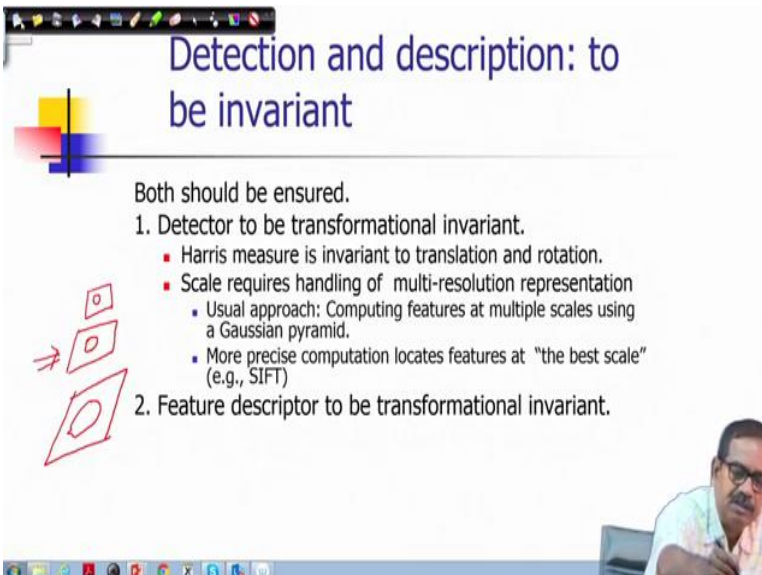
Just to explain a little bit about this resolution representation, what we can consider? Suppose you have an image and with an object this and it is given in its say original resolutions; that means, the camera resolutions the number of pixels whatever you have got in the sensors and the spatial resolution (Refer Time: 07:06) that particular number of pixels that gives you the; to gives you the highest resolution given that imaging.

But then what you can do, you can sub-sample you can down sample this pixels and get the smaller size of images. So, and in some cases even this particular it can could be so small that that this may appear like a dot. So, you can see that depending upon different resolution the structured structural information will vary. So, it may it may not retain the same structural information and sometimes the larger objects with when it has a very high resolution but once we increase the high resolution say you have a very very tiny objects. Now, in this resolution it could be detected whereas in this resolution this would be lost.

But you have a very large object in an image now with using local measures to get the overall ideas of the shape would be difficult to comprehend difficult to analyze. But if I get a smaller resolution of this image, then even a smaller window will be able to capture this particular feature then the local measure corresponding to that particular representation should give a higher value. So, that is the idea of having multi-resolution representation. So, you vary the resolution but highest resolution that is already constrained by the imaging system.

But only thing you what you can do you can get lower resolution versions and try to see that you know some of the structures in those lower resolutions are becoming more prominent and easily detective. So, this idea has been used to in particular this feature detection.

(Refer Slide Time: 09:27)



Detection and description: to be invariant

Both should be ensured.

1. Detector to be transformational invariant.
 - Harris measure is invariant to translation and rotation.
 - Scale requires handling of multi-resolution representation
 - Usual approach: Computing features at multiple scales using a Gaussian pyramid.
 - More precise computation locates features at "the best scale" (e.g., SIFT)
2. Feature descriptor to be transformational invariant.

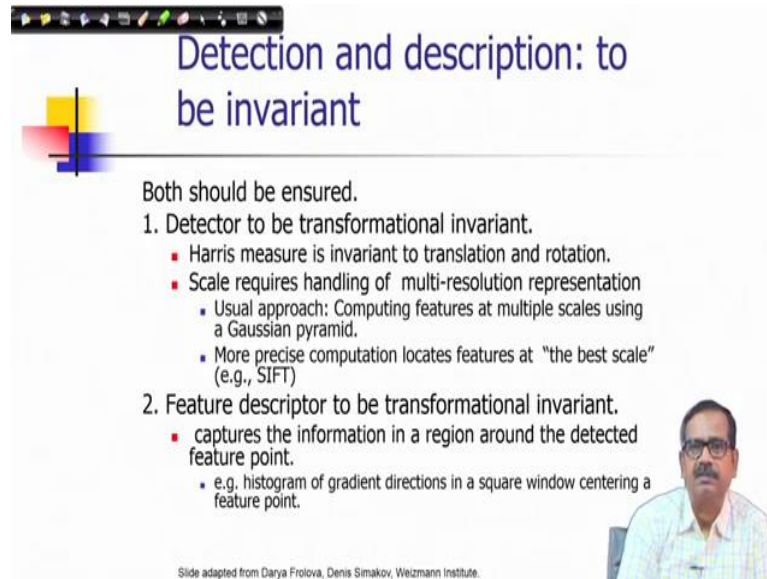
So, this is an approach, so you can compute features at multiple scales, you can use a Gaussian pyramid which means; you smoothen the image using some Gaussian mask varying standard deviation. So, scale would be higher iteratively; iteratively scale will be increasing and also you can down sample the image and you can get a pyramidal structure of that representation you have an image. So, if I show it in this form say let me show the image it is a bottom resolution higher resolution.

In the next version after smoothing this image using a Gaussian mask and such sampling you get this resolution you use further smoothing; that means, effectively you are using a very larger mask over this image. So, you're smoothing this image and you are getting coarser distribution and sub-sampling it. So, you get a next level of resolution.

So, this kind of representation is useful. Since the shapes looks like a pyramid if I place them if I stack those images in this particle vertical order it's just a visualization. So, that is a very popular term which is used for this multi resolution representation we call it a representation following a Gaussian pyramid representation. Which means every image is convolved by a Gaussian mask and then subsampled and you get multiple representation.

So, for single image you get 'n' number of images of varying sizes by this process. So, this is what Gaussian pyramid representation is. And in fact, there is a very efficient and effective you know method by which you can compute the best scale for feature detection and that method I will be discussing in a method called sift method which is scale invariant feature transformation. So, we will be discussing then and. So, the basic idea is a feature descriptor should be transformation invariant.

(Refer Slide Time: 11:53)



Detection and description: to be invariant

Both should be ensured.

1. Detector to be transformational invariant.
 - Harris measure is invariant to translation and rotation.
 - Scale requires handling of multi-resolution representation
 - Usual approach: Computing features at multiple scales using a Gaussian pyramid.
 - More precise computation locates features at "the best scale" (e.g., SIFT)
2. Feature descriptor to be transformational invariant.
 - captures the information in a region around the detected feature point.
 - e.g. histogram of gradient directions in a square window centering a feature point.

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

And it captures the information in a region around the detected feature point that should be the property. For example, we can consider histogram of gradient directions in a square window centering a feature point. So, these are the two steps one is the detection which should be invariant to transformation including scale, the other one is a description. So, you should consider now description at that scale.

So, whatever local statistics you collect you should collect the image which has been transformed through that multi-resolution processing; that means, which has been convolved using a Gaussian mask of that scale and then consider the point which is been detected at that scale and consider the neighboring statistics at that scale. So, these are the two policies which are used in particular to get it transformation invariant scale invariant description.

(Refer Slide Time: 12:51)

Finding Keypoints – Scale, Location

- How do we choose scale?

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, it just explains what I wanted to show through the diagram you can see at different resolution the local descriptions they two vary at an appropriate resolution. The alphabet 'a' is visible, but if you look at very closely then 'a' gets missing. So, if there is any measure of this interestingness of a particular resolution, then there is an appropriate resolution in the middle where you get more interestingness and that is what is shown here hypothetically, it has been shown by this particular curve.

(Refer Slide Time: 13:31)

Scale Invariant Detection

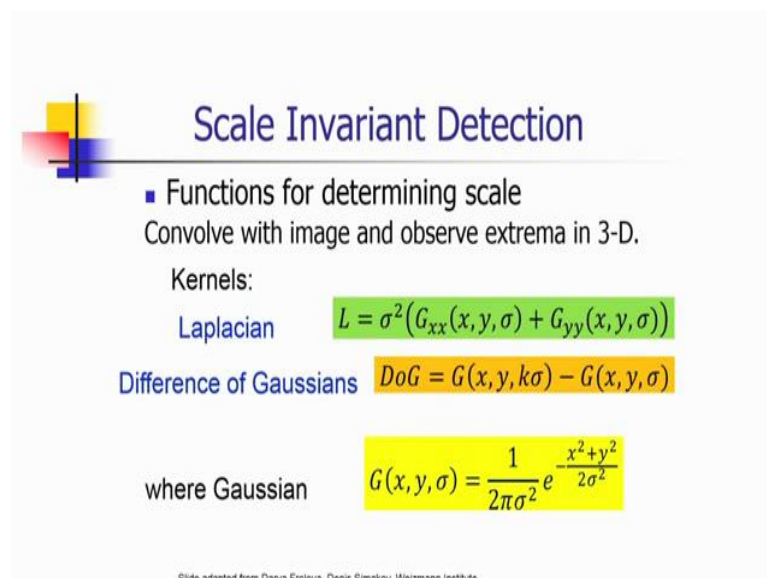
- Functions for determining scale
 - Convolve with image and observe extrema in 3-D.
 - Kernels:
 - Laplacian
 - Difference of Gaussians

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, for scale invariant detection one of the major task is that to determine the appropriate scale and for that we need to convolve with image. And since now we have a 3 dimensional space because not only the 2 dimensional special locations of say x and y direction, you have a direction along scale.

So, you observe the measures in both position and scale that would give the three dimensional space. And there are different kernels which we are going to define here like Laplacian kernels and difference of Gaussians which means kernels are here this is the masks as we defined earlier and which needs to be convolved with the image to give a measure and you would like to find out a local maximum of those measure.

(Refer Slide Time: 14:33)



Scale Invariant Detection

- Functions for determining scale
Convolve with image and observe extrema in 3-D.

Kernels:

Laplacian $L = \sigma^2(G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma))$

Difference of Gaussians $DoG = G(x, y, k\sigma) - G(x, y, \sigma)$

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

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, just to define it, first let us consider the definition of a Gaussian function. As you can see, this is a 2 dimensional Gaussian function and uniformly scaled along directions which means standard deviations is uniform in all directions if particular in two principal directions x and y directions and this is the you know Gaussian function and this is a continuous function.

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

So, for a discrete processing you need to get a discrete representation of this function over a mask or over a window. Say if I choose a mask size of say 10 x 10, then that fixes you have to find out considering the center of the mask as the origin you get the functional values in other locations.

And the mask size depends upon the values of sigma. So, the one of the criteria could be that it should be say $2\sqrt{2}\sigma$ sigma which is a very large mask size. So, this is a definition of the Laplacian mask and we have you have defined using the Gaussian function.

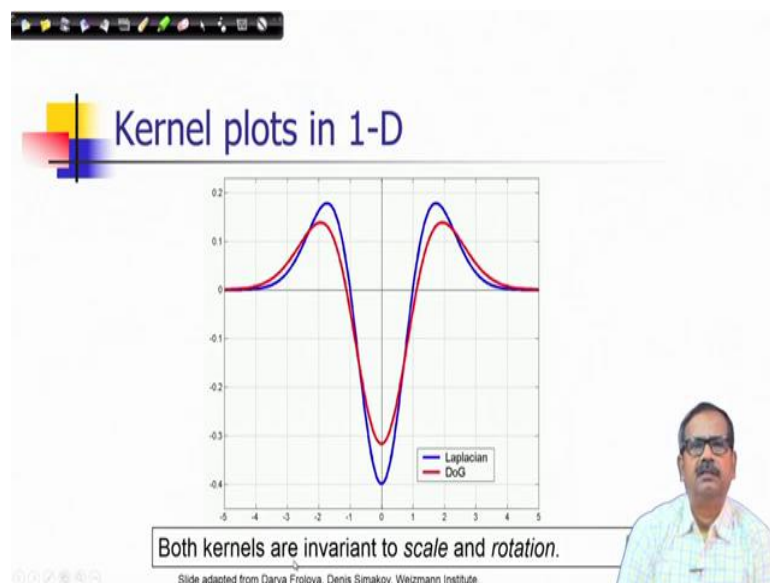
$$L = \sigma^2(G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma))$$

As you can see that these are the second derivatives of those Gaussian functions it's some of the second derivatives along x direction and y direction and this is normalized by multiplying with sigma squared to make it scale invariant description. And difference of Gaussian this descriptor is defined in this fashion

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

It is just from the nomenclature itself. It is understood that it's mask which is defined from the difference of two Gaussian functions of two different scales; one scale is $k\sigma$ the other scale is σ .

(Refer Slide Time: 16:25)



These are the shape of the kernel; that means, shape of those functions in 1-D you have shown and no any 2-D you just rotate it along the axis of symmetry in the center and; that means, about y axis if you rotate, then you will get the 3 dimensional mask. That means, it's a function of 2 dimensional space, but we will get the values in a represents as a 3 dimensional representation on that particular function. So, you can see from this particular plot that both difference of Gaussian and Laplacian they are quite similar.

(Refer Slide Time: 17:13)

Relationship between LoG and DoG operator

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G \Rightarrow \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2}$$

$$\sigma \nabla^2 G = \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

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

The factor $(k-1)$ is kept constant across scales.
 → does not influence extrema locations.

They are quite similar in fact, mathematically also one can show you can perform these operations; that means, take the Gaussian functions take the partial derivative along sigma, then you can show that that is equal to sigma into this Laplacian of G, Laplacian of that Gaussian mask.

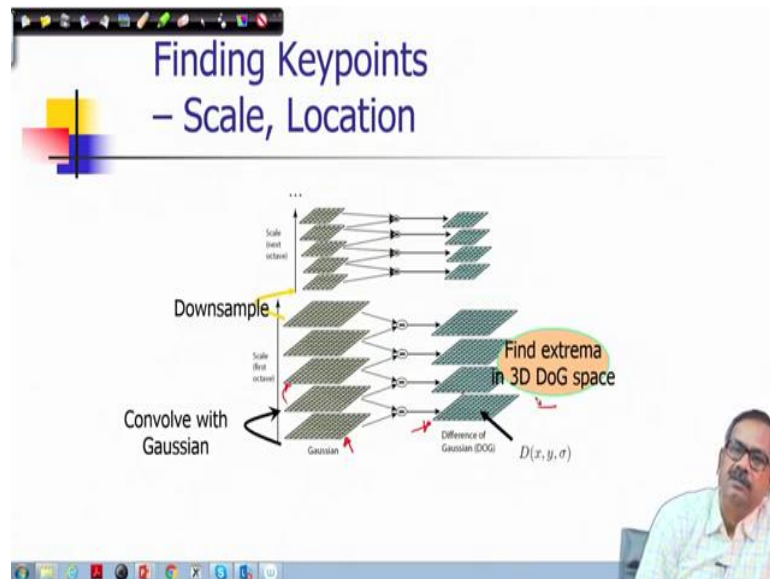
$$\frac{\partial G}{\partial x} = \sigma \Delta^2 G = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2}$$

So, this can be shown in this problem and you can see that the Gaussian difference of Gaussian mask is proportional to the corresponding Laplacian mask and this factor is $(k-1)$ is kept constant across scales.

$$G(x, y, k\sigma) - G(x, y, \sigma) = (k - 1)\sigma^2 \Delta^2 G$$

So, it does not influence extreme locations.

(Refer Slide Time: 18:17)



So, figuratively it is showing how this computation proceed. So, you have to convolve with a series of Gaussian masks and then at different layers sometimes you have to down sample it and produce that 3 dimensional functional distributions in a three dimensional space and then you have to get the local extreme.

So, that is what you will be considering. So, in this particular particular figure it has been shown the order of computation. So, this is the original image, you perform the Gaussian mask you perform the Gaussian operation. So, this is a Gaussian and then you get the subtraction that is a difference of Gaussian. So, this is the first representation of difference of Gaussian. Again you smooth using the Gaussian image Gaussian convolution subtract this one from this one. So, you get the second layer of difference of Gaussian images.

So, in this way you are producing the representation in a 3 dimensional space of scale and positions and then you have to find out the extrema in 3 dimensional DoG space or Difference of Gaussian space.

(Refer Slide Time: 19:37)



Scale Invariant Detectors

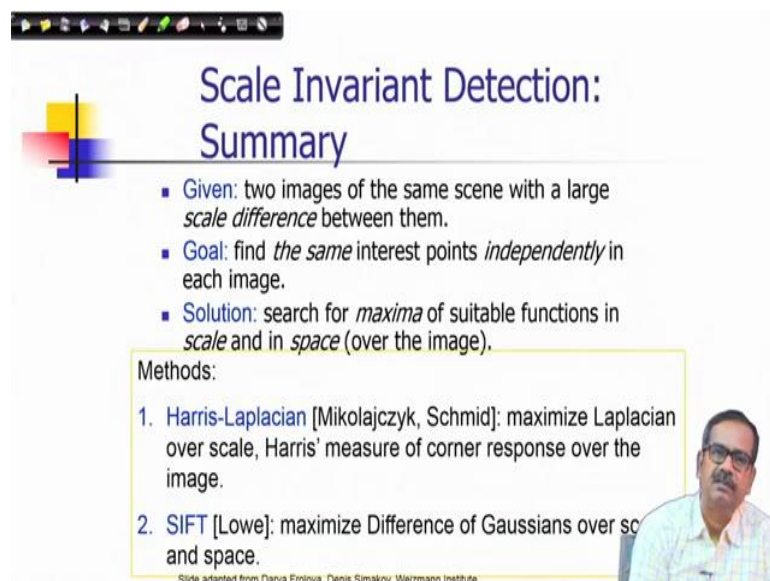
- **Harris-Laplacian** (Mikolajczyk & Schmid)
 - Apply Laplacian operation with varying scale.
 - Get local maxima of Harris corner response in space and scale.
- **SIFT** (Lowe)
 - Find local maximum Difference of Gaussians in space and scale

K. Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001.

D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints" 2004

So, just to summarize that we have discussed about two different kinds of detectors the difference of Gaussian detector is used in this sift descriptor which I will be discussing next and which has been proposed by David Lowe which has been found to be very popular. And also Harris Laplacian; that means, Harris operator with applied over the Laplacian pyramid representations of the particular with varying scale and then you get local maxima varies corner response in that space and scale that would also give you the transformation invariant. So, these are the two major you know detectors which are popular in particular in the literature.

(Refer Slide Time: 20:27)



Scale Invariant Detection: Summary

- **Given:** two images of the same scene with a large *scale difference* between them.
- **Goal:** find *the same* interest points *independently* in each image.
- **Solution:** search for *maxima* of suitable functions in *scale* and in *space* (over the image).

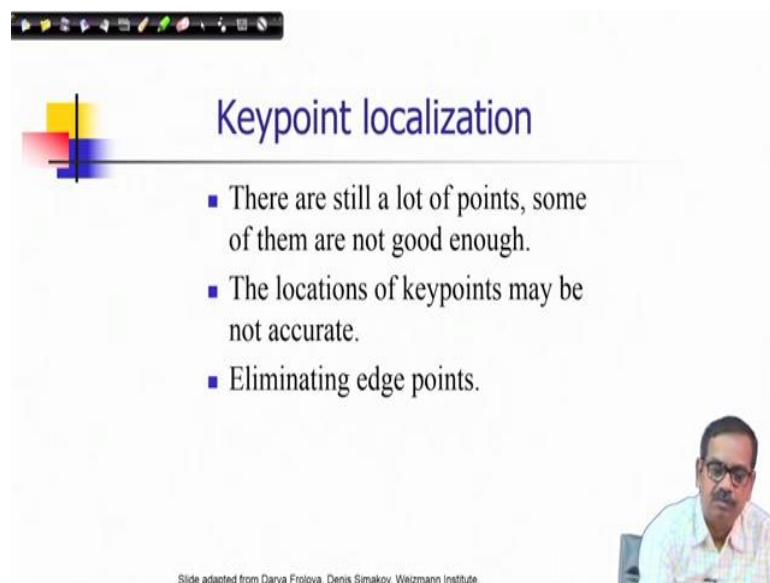
Methods:

1. **Harris-Laplacian** [Mikolajczyk, Schmid]: maximize Laplacian over scale, Harris' measure of corner response over the image.
2. **SIFT** [Lowe]: maximize Difference of Gaussians over scale and space.

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, just to summarize the scale invariant detection that given two images of the same scene with a large scale difference between them, we have to find the same interest points independently in each image and so the solution is that you have to search for maxima of suitable functions in scale and space over the image. And these are the two methods which I mentioned one is Harris Laplacian and this it maximizes the Laplacians over scale and Harris is measure of corner responses that has to be used in those Laplacian over scale representation and then the sift is maximize the difference of Gaussians over scale and space.

(Refer Slide Time: 21:17)



The slide features a title 'Keypoint localization' in blue text. To the left of the title is a graphic consisting of a vertical yellow bar, a horizontal red bar, and a horizontal blue bar, all intersecting at a central point. Below the title is a list of three bullet points, each starting with a blue square. In the bottom right corner of the slide, there is a small video inset showing a man with glasses and a mustache, wearing a light-colored shirt, speaking. At the very bottom of the slide, there is a small line of text: 'Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.'

Keypoint localization

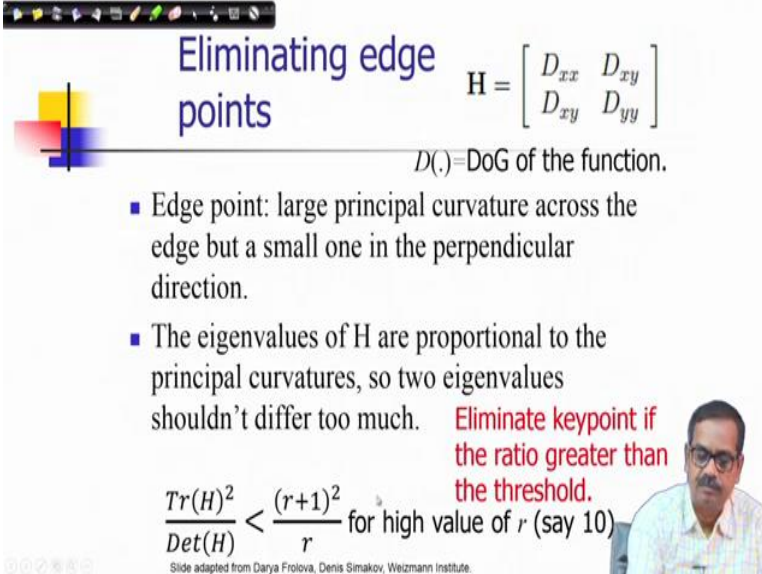
- There are still a lot of points, some of them are not good enough.
- The locations of keypoints may be not accurate.
- Eliminating edge points.

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

Now, the thing is that what you get out of this process is you get lot of key points. Now, some of this key points are not so important and some of them may not be structurally very robust because a small disturbance can disturb them.

So, we would like to get only those key for a points which are more robust to the transformation and which has a very precise location locations are to be very precisely defined there. And for example, a many key points will lie on edges and as we have discussed that, corners are more robust than edges. So, even some edge points can give a very high response and can be a local maxima, but we need to eliminate those edge points.

(Refer Slide Time: 22:11)



Eliminating edge points

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

$D(.) = \text{DoG of the function.}$

- Edge point: large principal curvature across the edge but a small one in the perpendicular direction.
- The eigenvalues of H are proportional to the principal curvatures, so two eigenvalues shouldn't differ too much. **Eliminate keypoint if the ratio greater than the threshold.**

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r+1)^2}{r} \text{ for high value of } r \text{ (say 10)}$$

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, there are certain operations like you can perform over the key points you can apply this particular operations. For example, if you apply this particular Laplacian operator that is edge operator what you can see, if these are the double derivatives over the difference of Gaussian functional values and it is expected that this would give you the curvature value.

So, the Eigen values of this particular matrix.

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

will be giving the principal curvatures and it is expected that for edge points it would be large across the edge, but a small one in the perpendicular directions. So, both the curvatures should be also very large. So, this is one characterization of edge point by which you can eliminate those edge points.

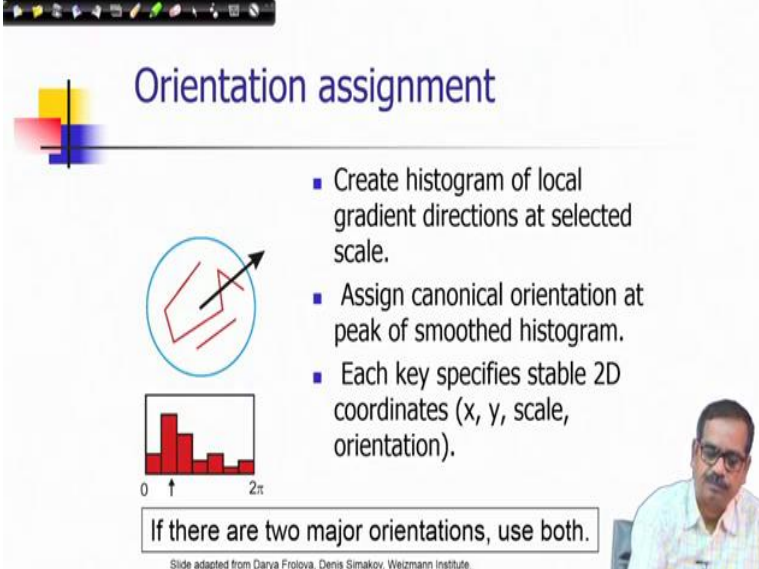
$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r + 1)^2}{2} \text{ for high value of } r \text{ (say 10)}$$

S so you compute the eigenvalues and eigenvalue should be large they should not differ much too much.

And this is how this computation is carried out and this is equivalent once again to you know equivalently computed by computing traces of this matrix H which has been defined

here, this H is different than what we discussed for Harris operator because these are all double derivatives of the difference of Gaussian functions as you can see from the definition. So, the ratios of square of trace and determinant it should be very high then only for high value of 'r'. So, it should be less than this and then you should we should take this value. So, eliminate key point if the ratio greater than this threshold it occurs if the ratio is greater than this then those key points are eliminated.

(Refer Slide Time: 24:19)



Orientation assignment

- Create histogram of local gradient directions at selected scale.
- Assign canonical orientation at peak of smoothed histogram.
- Each key specifies stable 2D coordinates (x , y , scale, orientation).

If there are two major orientations, use both.

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

So, now the question is that how do you characterize a key point? So, there are some local statistics that we need to consider, one of the attribute that we would be considering that what the major orientation around that neighborhood is. So, we can assign that orientation because if you get the major orientation, then the local statistics could be made orientation independent you can perform a transformation. So, that your reference axis is aligned to that major orientation or major oriented directions and then you can aggregate those local statistics by performing those transformation that is how you can make it rotational invariant.

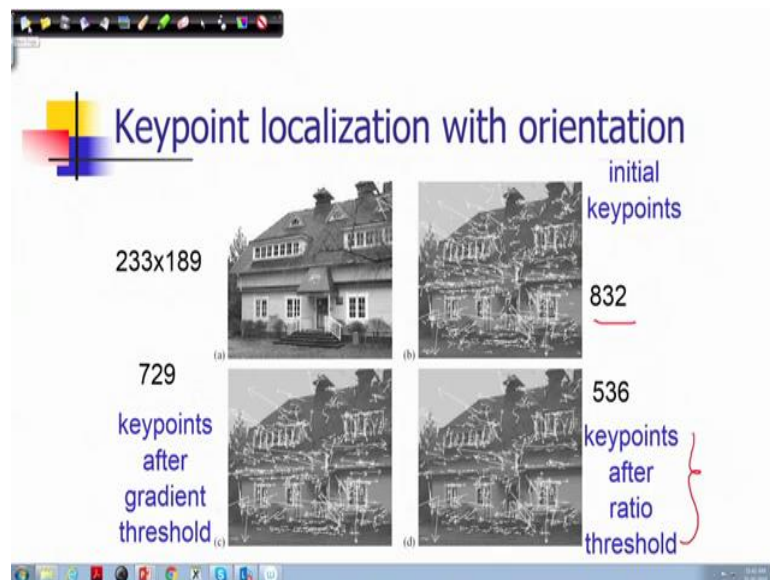
So, computation of orientation is important. So, what we can do that, in that case is locally you can get gradient directions around its neighborhood and then you can compute a histograms and bin them. You can see that in this particular example this binning of the directions are shown by this angles because this directions can be considered with respect to some x axis with respect to the reference x axis the angle what is formed by this direction

that is what is of our interest. So, you can discretize this range of this angles varying from 0 to 2π into some intervals and then put those directions into one of those bin.

So, that is what is known as binning of these directions in a histogram and then find out know which is the prominent one out of this discrete options and we can assign that directions to that particular feature vector to that key point actually. So, assign canonical orientation at peak of smoothed histogram. So, even you can perform smoothing of this histogram and then you compute the peak of that particular functional value and that peak will give you the orientation.

So, in this way a key point is described by its position scale and orientation because as I mentioned that scale is determined where you get also the local maxima in position and scale that is of the scale and positions have obtained and then orientation is computed in this in this fashion. There could be some situations where you can have two major orientation. So, you may have to use both; that means, multiple descriptor description of the same key point.

(Refer Slide Time: 27:05)

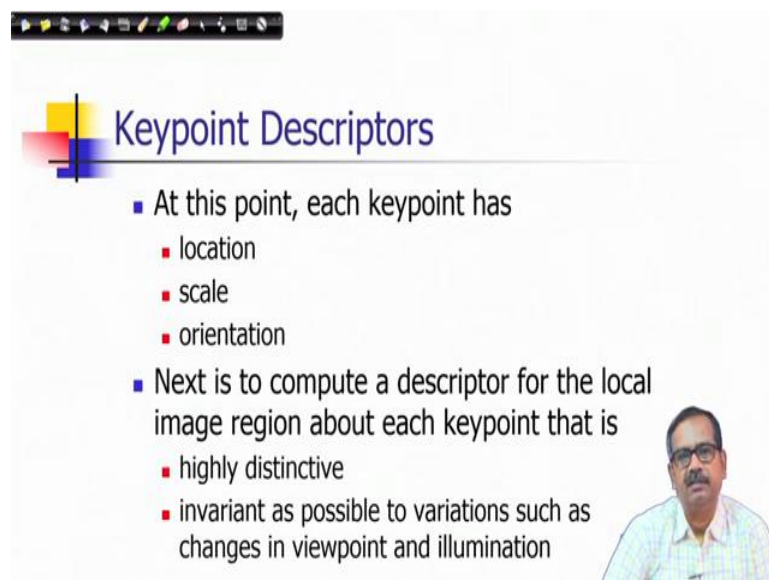


So, these are some examples which are again taken from this slides. In fact, it is also in the book by Zisserman. So, you can see that I mean Zisserman and Hartley there are multiple authors. So, there are there are some examples that you have key points after gradient threshold and key points after ratio threshold this example shows that how key points are you know reduced using those different kinds of processing.

So, initially you have this many number of key points on the same image 832 and it there are different orientations it shows a orientations and also you know the scale is associated with a position is associated with position has been shown scale is difficult to show in this particular diagram.

And then after performing a gradient threshold you can reduced to a number 729 and then after performing that ratio threshold which means using this curvature analysis of the Hessian matrix that is called Hessian matrix of you know difference of Gaussian function and from there you can perform this ratio threshold, you can get reduce more number of key points.

(Refer Slide Time: 28:33)



The slide is titled "Keypoint Descriptors" and features a list of bullet points. The first bullet point states that each keypoint has location, scale, and orientation. The second bullet point states that the next step is to compute a descriptor for the local image region about each keypoint that is highly distinctive and invariant to variations such as changes in viewpoint and illumination. A small inset photo of a man with glasses and a mustache is visible in the bottom right corner of the slide.

- At this point, each keypoint has
 - location
 - scale
 - orientation
- Next is to compute a descriptor for the local image region about each keypoint that is
 - highly distinctive
 - invariant as possible to variations such as changes in viewpoint and illumination

So, this is a summary of a key point characterizations that it has a location, it has a scale it has an orientation. So, next we need to discuss that how to compute a descriptor for the local image region about each key point and that should be very highly distinctive and also it should be invariant as possible as for the variations such as changes in view point and elimination. So, we will continue this discussion in the next lecture.

Thank you very much for your listening.

Keywords: Feature matching, detectors, descriptors, keypoints