Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-54

So, it is like this right, so you are here and you have taken a 3 cross 3 neighborhood, 3 cross 3 on top and then 3 cross 3 at the bottom right and you are kind of here and you are searching for searching above, searching below and searching across right. And within that you see 26 kind of you know neighborhood whatever turns out to be the maximum you take to be the to be the interest point. Now once you have an interest point again that you have to somehow follow you know roughly what we did earlier, what is the next thing that we do? We give an orientation assignment right, so once this is done right what you need to do is do the orientation assignment. This is done a slightly I think this is also roughly very similar to what we did before but then only difference is it uses the Haar wavelet okay earlier because this has to do everything with a with a kind of box filter right. So, it uses a Haar wavelet okay this is done using a Haar wavelet, so Haar wavelet looks like this again right the scale the size of this Haar wavelet right will be a function of the scale at which you are applying okay it is not a fixed size. So, you have like you know 2 of them, one to sense the vertical and then one other to sense the horizontal change, so this will be like minus 1, 1 and then what about minus 1, 1 and yeah so the size is scale dependent.

I think there is even some notion of what is the size I think, so the size is 6S wait a minute then a 4S, so this is taken as 4, so S is your σ okay or 4S, S is a scale at which at which you found the extrema right when at whatever scale you found it. So whatever is that σ 4 times that σ approximation of that is what will be this Haar wavelet okay that will be the size and always remember the Haar wavelet is such that its area and all will be 0 and so on it is actually that will sum to 0 its weights and all and that is how it is again chosen okay. So it is a function of the scale then again right I mean what you need to do is I mean you are sitting at a point and then you want to know kind of some orientation assignment you want to do what we were doing earlier it were now we are trying to find out the right we would find the strength of the gradient then we will know the orientation will actually bin them right I mean something similar to that except that there are some minor changes. So within a circular neighbor okay let us not be too specific, so within a neighborhood of 6S around the interest point okay these are all okay 6S around the around within a neighborhood around the interest point the responses are weighted that means you apply the Haar wavelet right let us say you get an X response and then right you get a you get a response along Y the responses from the Haar wavelet the Haar wavelets are weighted responses let us say to the Haar wavelets the response to the Haar wavelets are weighted with a Gaussian all these are hyper parameters by the way with a Gaussian σ equal to 2S.

So this is only to weight so it is like saying that if you are farther away in that 6 σ 6S neighborhood right then you are whatever is it is right that you actually contribute your contribution should be less if you are farther away from the interest point. So it is like with a Gaussian 2S so maybe the guys outside you know will not even contribute far no which are farther away from the center okay so the response are actually weighted with this one Gaussian σ equal to 2S then the vector sums the vector sums so by the vector sums right what they actually mean is you have the X response you have the Y response right and then of course you know you also you also have the angle right with you therefore what you do is you know you have a 30 degree bin I mean that each bin of bin of you see 30 degrees and all those rate which fall in that is the orientation like 30, 60, 90 right. So wherever you have an orientation that falls in a particular bin you just add up the X responses weighted of course by the Gaussian then add up the Y responses and rate keep them all in that in one bin and then go to the next bin find out what all are falling within that add up all the so do a vector sum so that you have like an X vector sum and then a Y vector sum within a bin and then and then right if you add that then you have a sum vector right I mean a resultant vector and whichever resultant vector turns out to be the longest rate or has the highest length or the maximum strength right there way if you think about it as gradient value then whichever this all very similar to what we did earlier there you know we were just doing I think gx square + gy square or something here it is like a vector sum that is the only difference and whichever gives you the longest length you believe that that is the orientation that should be attributed to that interest point right. Okay so the vector sums of all the weighted gradients of all the weighted gradients all the weighted gradients in the 6S neighborhood within 30 degree bins are actually computed and the orientation of the interest point is actually defined as the orientation of the interest point is defined as the global maximum or the or the or is defined as the orientation corresponding to the longest vector corresponding to the longest sum vector. Sum vector means the response is added along X and Y okay that is the orientation assignment very similar to shift but then here everything is done with all this again right you can use integral images so nowhere do they kind of except for this Gaussian weighting and all that that of course that is still done using traditional but otherwise right wherever possible the approximation is always using like a box filter.

Then a descriptor right self descriptor now that right now that you know the interest points you need a descriptor so again I will just tell you what it is so choose a window of size 20S again right these are all coming from the paper okay choose a window of size 20S around the interest point and align it along the dominant orientation. So there right if you saw we were doing something like a difference we used to compute the difference right with respect to this reference orientation here alignment there is there is rotate the patch itself by that angle and align it along so this window of size 20S right which are actually picking right that itself is actually rotated now with respect to this orientation which is the most dominant and align it along the along the dominant orientation of the of the key point of the key point. Of course we have not talked about how to do the you know rotation and all but those are all straightforward things okay and this region is then split into 4 cross 4 regions so this region is split just like we did there right into 4 cross 4 sub region so whatever be the depending upon the scale right all the sub region may have lots of pixels and so on so you have this 20S neighborhood right and around this we are kind of dividing into 4 cross 4 regions right 16 of them sub regions and this is done to keep the keep important spatial information right this is again the same way as in SIFTED we said that we will divide there also we did something like 4 cross 4 right and then we kind of kept the spatial information by actually taking we had an we had an we had an whatever 8 cross 1 vector coming from each region and then we aligned them right there were 16 of them so we got a 128 vector. So that 128 vector but that is a there is a spatial alignment that you keep right it is like you go in some order and you keep that because that is a spatial alignment which you want actually preserve so that is happening even here so the region is split out right to keep the spatial information to actually keep this the spatial information. Then then HAR for each sub region for each sub region HAR wavelet responses dx and dy are computed at 5 cross 5 regular regularly spaced sample points again these are all you know as specified in the paper okay at regularly spaced intervals spaced intervals.

So it is like you compute dx and dy within okay this is all this is all within this sub region and again right these are again there is a Gaussian weighting these are Gaussian weighted that means you will have a Gaussian that is kind of dying off right at the at the edges so these are again Gaussian weighted okay not not not with respect to this center sorry it is with respect to this center the center of the the main interest point these are Gaussian weighted this is a Gaussian weighted with respect to the interest point okay not not not I mean within the sub region with respect to the interest point. I mean again it is about how far away you are from the actual interest point not like whether you are at the center of that sub region okay then the dx and dy are all summed up within a sub region are summed up I mean it is actually a vector sum right are summed up within a sub region and that along with magnitude of dx and dy it becomes becomes becomes one vector so you have like for one sub region that you have dx which is a summed up value right computed within the sub region taking 5 cross 5 patches and then applying a Haar wavelet on that and then Gaussian weighting it with respect to the key point location then you have dx then you have dy then you have mod dx then you have mod dy okay that becomes a 4 cross 1 vector right for one sub region and we got 16 of them right. So since there are since there are 16 sub regions there are 16 such sub regions the dimension of the feature vector turns out to be 64 cross 1 feature this one a descriptor is 64 cross 1 and this 64 cross 1 is what now right you can use in order to do your matching you can say detect and match and all the invariances are all that we talked about right with respect to say all hold here hold right here may know here also I mean again that you are taking you know this one Laplacian therefore illumination invariance right to a certain extent exist and this is also normalized okay and this vector is also normalized to actually unit length just like just as we did there normalized to unit length so as to take care of take care of any sort of a contrast right non uniform illumination as well as uniform illumination to take care of that this is normalized to unit length and that is a rotation invariance is there the scale invariance is also there and therefore it is on a sense I mean whatever we saw in shift is all there except that this is easily a 5x speedup okay. So doing all of this you still get a 5x speedup over shift right because of the approximations right that were actually made using box filters. See there is one more thing okay there is something called the histogram of oriented gradient okay it is called the hog but that is very similar to whatever we have done already in terms of the orientation, binning and all that except that it is actually a dense filter so actually dense kind of you know I mean you get dense points whereas here you get sparse point that is only difference but otherwise the idea and all that is very similar so I thought there is no point you know spending time on that because any of you can just read it up now that you have so much background in this feature detection and all that should be just very easy right.

So I thought I will not spend time on hog leave it to you to okay read it up it is very simple now that I mean it is all based upon whatever you already it is much simpler than that in fact it was originally used by this I mean it was actually proposed by 2 people okay Dalal and Triggs this is one of them is an Indian actually so they actually did it for you know for actually human detection you know people detection and therefore if you look at the size and all right it will be like 11 aspect ratio which is like 1 is to 2 because that is all tailored towards humans but later it is also been used for other things. So it is very simple to understand okay so far less complicated than these other descriptors you know but it is more dense okay so I will leave that to you and from next class right we will start geometry okay.