Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-73

Okay, so let us go ahead and complete what we started last time, which is this factorization method and I had come up to the point, right, where I think we had something like W is equal to what was that Ms + T, is that what we had used, yeah, M into S + T. T had translations in it and so you had something like F number of frames, right, so you had like, what is it, F number of frames, so you have like 2F cross the number of points is Q, right and M had 2 rows for every frame, therefore this must be like 2F cross and 3, right, every row has 3 coordinates, it is actually a rotation and this S had 3 and then there were, so this was a structure, right, so this had, what was it, 3 cross Q, right, you had Q scene points and therefore, right, this guy will be like 2F cross Q and this was a measurement which meant that along a column, you had a correspondence across F frames, where you had X and Y coordinates and you had Q such columns, right, that was the understanding. Now let us actually define a W ~, okay, first of all, before that, what will be the rank of W, what can it be, rank of W, less than or equal of course, we cannot say what is the actual rank, what will it be like, less than or equal to, what can be the rank of S, M maximum, 3, what can be the rank of S, 3, okay, what can be the rank of M into S, no, how do you find that, I mean if I have 2 matrices A into A dot B, what is rank of A dot B in terms of rank, individual ranks of A and B, it is minimum of rank A, rank B, okay, so we do not know, I mean S we are saying is less than or equal to 3, rank of M is less than or equal to 3 but which one is smaller and all, we do not know, so and T has a rank, rank is 1, it was the same column getting repeated, no, can you guys not recollect, what was it like, T1, T1, T1, T1, right, T21, T21, T21, it was just the same column getting repeated, so the rank of T is actually less than or equal to 1, okay. Therefore, rank of W, what is the rank of A + B, less than or equal to rank of A + rank of B, therefore rank of W is less than or equal to 4, okay, rank of T is less than or equal to 1, rank of M is less than or equal to 3 and therefore rank of W can be less than or equal to 4 but now, right we are going to actually nullify the translation and make it rank 3. So, how do we do that is you know, let us actually define a W ~ which is let us say W into, what do we do, I think we do something like I, an identity which is Q cross Q - 1 by Q all 1's, this is also a Q cross Q matrix of course, what does this do, if you do a matrix like this, what does it mean, what is the action of a matrix like this, I all diagonals, 1's along the diagonal and then 1, 1, 1 along a row and I am not doing Q square, you could notice that I am doing 1 by Q, when do you do this, I mean let us not go into signal processing and all, you have a data and this is called the data centering and it is like, yeah, we are moving, just mean centering, right, mean centred. So, all that this will do is this will give you a mean centred, so let us say W, mean centred W is actually W ~, okay, now you replace W by W by MS + T, right that is coming from here and then you have the same guy, I Q cross Q - 1 by Q, this one's all 1's guys.

So, now if you expand this, right, this will be like MS and all these dimensions are all will match, okay, there is no problem, I Q cross Q - 1 by Q + T, T is a matrix by the way, this is not a translational vector, - 1 by Q, so what will be the second term, what will be the second term, second term is actually 0, not because I - 1 by Q is 0, not because of that but then because T is all the same column repeated. So, look at the entries of T, right, it was like T11, T11, T11, the first row, then next was what, T21, T21, T21, so if you do mean mean is T11 itself, right, if you do that you will get all entry 0, so every entry, right, so this mean centred means every entry is subtracted from the actually mean value of that row, so mean centred means what, I mean every element of, so let me write this down, right, I mean if you have still not got in it, every element of W is subtracted, no not subtract, I mean from every element of W, the other way round, otherwise you will get a - sign, from every element of W you subtract the mean of the row, of that row, you subtract the mean value of that row, okay. So, in this case, as you can imagine the mean is the same as every element, it is all taken from T11, T11, therefore this goes to 0 because of the nature of, this is going to a 0 matrix of whatever size, right, what will be its size now, 2F cross Q, okay, does not matter, it is anyway 0, so due to the nature of T not because of anything else, due to the nature of T, T being, right, what it is and therefore, and now this one, okay, this is like what, S into I Q cross Q - 1, that is like mean centred structure now, right, whatever we did for W ~, now we are doing it, it is just invariably it is coming and sitting on S now, therefore this we can write as M into let us say, S hat, where I absorb this I Q Q into S, okay and therefore this I call as mean centred structure. So, I have W ~, right, is equal to this, now if you look at the size of M, it is 2F cross 3 and S hat is 3 cross Q, okay, now as always, right, we would want to sort of look at the SVD now of we say W ~, W ~, right, if you do an SVD, right, first of all what is M into S hat, so we just know, so I know, so M has, M has how many, this one 2F, right, 2F cross 3, right because each row is 3 cross 1 and S is like 3 cross Q, right, this is what you have now S hat.

Now, if you do an SVD, right, so W ~ will be like U diagonal V transpose, where this will be 2F cross 2F, this will be 2F cross Q and V will be Q cross Q, so that this becomes 2F cross Q, right, that is how it will be, no, but now, but then the point is, okay, what is the, what is the, what is the rank of this guy, rank of W ~, see if you notice rank of W was less than or equal to 4, but that had a translational component sitting whose rank was itself less than or equal to 1, so W ~ has nullified the translation also, therefore this will be, right, minimum of the rank of W, rank of S hat, right, so rank of S hat can be a max of 3, rank of M can be a max of 3 and therefore rank of W hat cannot exceed 3, right. I mean, why we are interested in this because this D, what is the point in having a D that is that big, right, when the rank of W ~ is only, can at most be 3, correct, so now what you can do is you can actually simplify this, right, I mean you do not have to have such a kind of, such a kind of complicated big D, right, when there is, when actually hardly rank of W ~ is like less than or equal to 3, therefore, right, what you can do is, so you can actually pick the, so what would you do, right, so you pick the, I mean let us kind of say define something like define, what are we defining, okay, U ‘, okay, U ‘ to be the first 3 columns of U and all this is assumed to be arranged, okay, right, I mean, you know, arranged in the order that, that you have a largest singular value down to the smallest and so on, all that we are assuming, so, right, define you had to be the first 3 columns of U, then take D hat to be the, to be the, to be a principle 3 cross 3 block, 3 cross 3 block of D and V ‘ whatever, right, let us say, what is it, the notation that I am using is okay, V ‘ transpose to be, you can either say the, I mean 3 rows of V transpose or write equivalently 3 columns of V, right, let us say in terms of column to be the first 3 columns of V, columns of V, right. And therefore, what you can do is, you can effectively write W ~ which is actually 2F cross Q, we can write this as U ‘, not U ‘, I mean, yeah, U ‘, right, it is not a transpose by the way, these are some other matrix, okay, U ‘ whose size is 2F cross 3 now and D ‘ is 3 cross 3 and your V ‘ transpose, right, as we are writing is 3 cross Q, okay. And this you can further kind of, you know, split as U ‘ √ D ‘, √ D, I mean, √ D ‘ you know, right, just a matter of taking the square √ of all the diagonal values, V ‘ transpose, okay. And therefore, this you can call as M ~ and this you can call as your, you see S ~ if you wish.

And therefore, right, what you have kind of, right, effectively achieved is, I mean, you have the observations with you which are the feature correspondences coming from across whatever F frames, no sorry, yeah, F frames and there are Q scene points and this observations, right, that you have, I mean, for let us say, you know, in this matrix W ~, you are able to split it up into decompose into actually rotations which is what M ~ contains rotations and then the scene structure, right, which is built in this S ~, right, this has the 3D scene in it, right, that has like X1, Y1, Z1 all the way up to Q, right, XQ, YQ, ZQ and that is what you have done and you will still be probably interested in the say, translations also because you need the pose, right, after all you are solving for the pose, you need rotations as well as translations. So this is the simplest thing to do is get back your, what do you say, right, get back your say, S hat, I mean, S ~, so you have your S, I mean, you have this mean subtracted thing here, no, that I showed earlier, so just go back there, right, so where was this mean subtracted thing, okay, so yeah, so from here, right, you have your S hat which is S into this mean subtracted, so you can actually get your S and once you have your S, you have your M and therefore, right, all that you need to do is you need to take your, I mean, you have your original W, right, so translations, okay, get S from say, S hat, it is easy because it is just that I – that operation and just take it on to the other side, then translation is simply the original matrix W – what you call MS, that is it, right, so you have your translations, you have your, see, your rotations and you have your C3D scene, okay, this is called the factorization method and still the point is that this is not the final call, right, in the sense that one, of course, we have assumed that, you know, that the model is orthographic which was a simplification that we made but the whole point was to convey how this is done, I mean, you could have done with actually, you know, this one, a perspective sort of a projection model but that would have been more involved, I thought it is good to sort of look at something that is more, that is kind of more easy to understand and of course, this is what is also normally taught, okay, this orthogonal deflection projection model is what people typically use and now you know how this is done but then, right, it does not mean that, it does not mean that, it means that you have, you know, sort of a dense 3D reconstruction, right, you do not have, you still have only a few points, right, for which you have this sparse feature correspondences, you have a 3D, you know, 3D points for that and then the key thing is you have the poses now but even the poses need not be accurate, right, because one of the things that we were assuming here is that our cover is a correspondences are all noise free and all which will not be the case, right, it can either have noise or it could be outright wrong, right, both can happen. So what is normally done is these are all taken as initial estimates only, okay, so you can actually use DLT what is called direct linear transform which is the other route, right, which is to kind of compute the fundamental matrix then compute the essential matrix then from the, from the, say, right, essential matrix you can go back and, and, you know, find your, find your, find your, what you call, R and T, right, from the essential matrix then you can actually triangulate and then you can scale your component, right, as you go from one path to another, so there are various ways in which you can arrive at the initial estimates but, but what is typically done in a, in a kind of structure from motion is that whatever initial estimate you have, those are still considered as initial estimates but the actual, the, the real thing that happens is what is called, what is called a bundle adjustment, okay, this is where, this is where all the, all the right errors and all are actually accounted for because your, say, correspondences need not be perfect, there could be noise in the way you have located the correspondences, there could be outliers in the sense that you have probably matched things wrongly because you are just depending on something which is just something like a sift or surf, right, which is actually giving you the, giving you the feature correspondences.