Indian Institute of Technology Kanpur

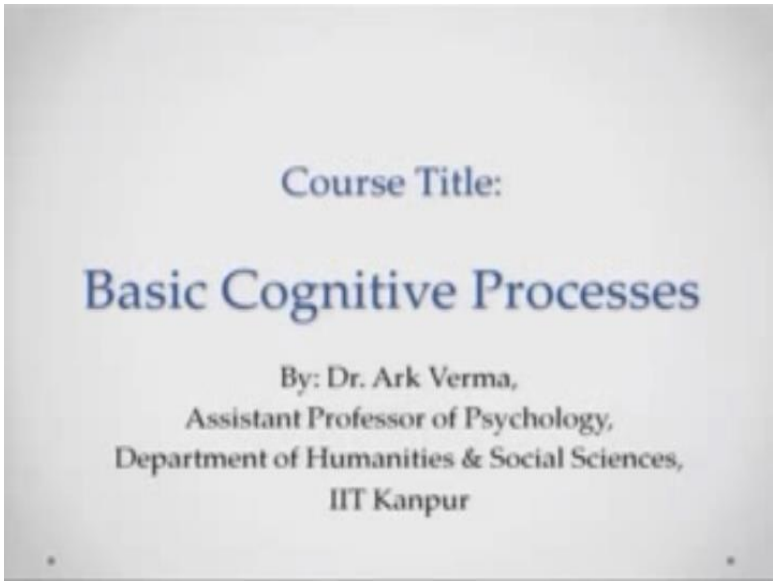
National Programme on Technology Enhanced Learning (NPTEL)

Course Title
Basic Cognitive Processes

Lecture-21
Theories of Object Recognition

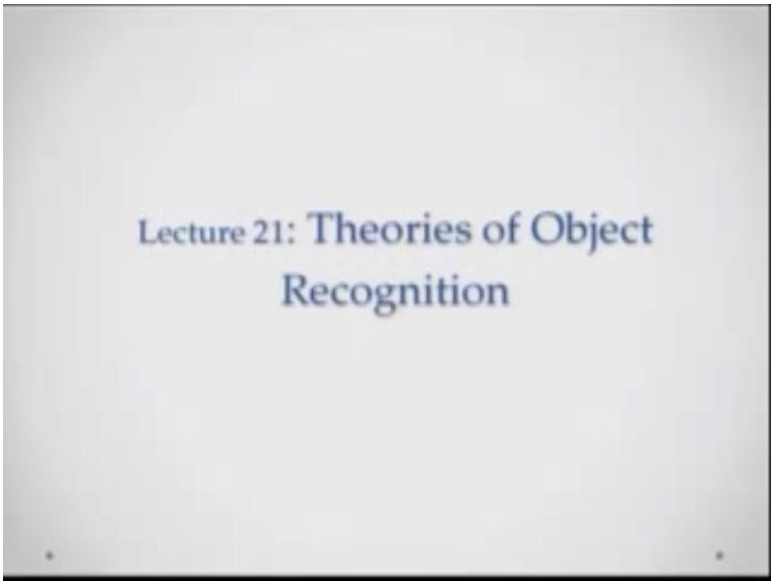
By
Prof. Ark Verma
Department of Humanities and Social Sciences
Indian Institute of Technology Kanpur

(Refer Slide Time: 00:17)



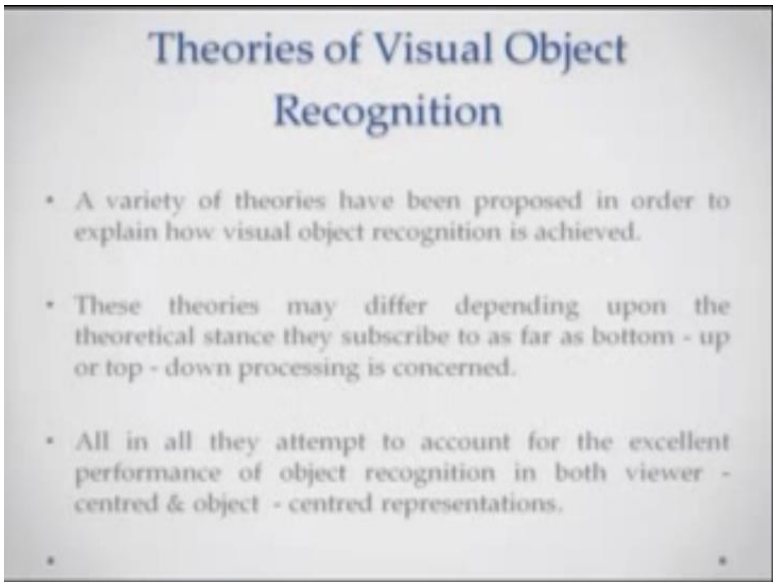
Hello and welcome to the lecture series on basic cognitive processes I am Dr. Ark Verma from IIT Kanpur.

(Refer Slide Time: 00:23)



Today we will talk about theories of object recognition you have seen till now that we have been talking about perception we have been talking about various theories of perception today we take a particular case and we see how we interact and how we understand and recognize objects in this external world now there have been a variety of theories that have been proposed in order to explain how visual recognition is achieved.

(Refer Slide Time: 00:48)



These theories may differ depending upon the territory stance they take say for example whether they are bottom-up theories or they are top-down theories you might know already using the previous lectures that bottom-up theories basically focus upon developing or using the information coming from the sensory experience to develop mental representations and top-down theory is basically a favored the theoretically stance that you know it is our memory and it is our experience in knowledge of the world.

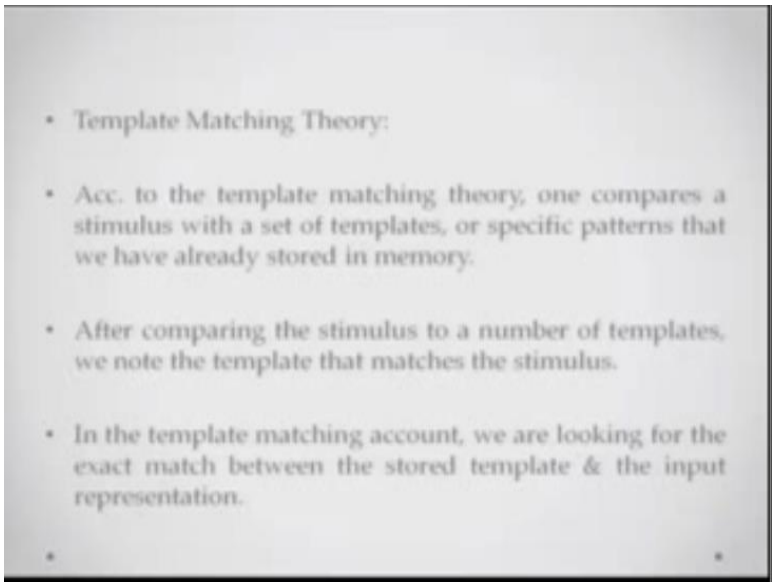
That helps us build maintain representations of the external world now we will just we're just adding a case there we are saying whether it is the sensory experience that leads us to form representations of the objects whether it is that information that helps us to recognize and interact with objects versus whether it is our memory and our knowledge of the world that helps us interact and recognize objects that is the kind of a difference.

You know we can talk about all in all in all basically the attempt of this object recognition theories is basically to be able to account for the excellent capacity of object recognition that we have the fact that we make errors very rarely and the fact that we do object recognition rather quickly there are two problems of object recognition we refer to this why we were talking about

Mark there are two kinds of representation possible if you are talking about a particular object one form of representation is if you are taking an object in variant view are you talking about the representation of the object which is an object centered representation versus.

If you are talking about viewer centered representation say for example I am looking at a flower pot you know standing on the window of my house how does that flower pot look if I move around versus what is the general perception of that flower pot which will not change irrespective of how I move around about in the room so these are some of the problems which are there in object recognition we try and see how these various theories have attempted to solve those problems.

(Refer Slide Time: 03:15)

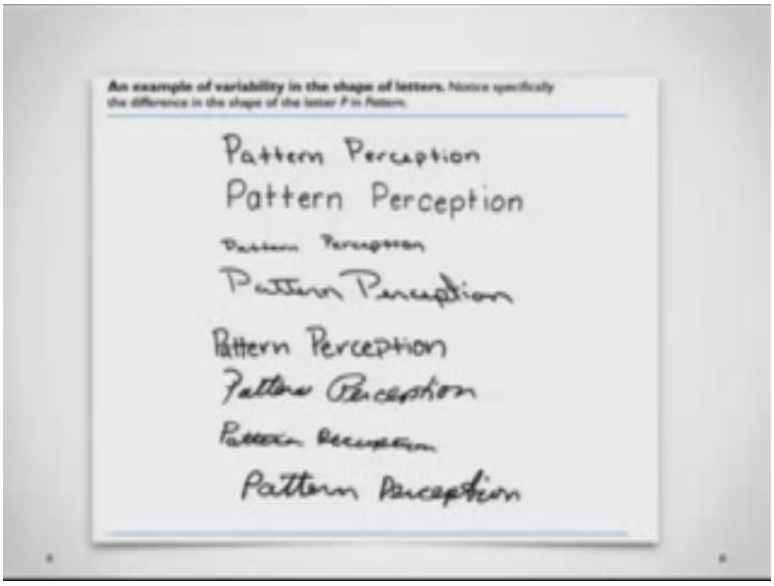


Now they have their variety of theories we will talk about that but one of the most basic theories of object recognition could be something like a template matching theory now our template matching Theory basically says that you know we compare the stimulus then we compare the sensory input with a set of templates that we already have okay so you might already have a template of how you know a particular ball looks like or how a particular you know toy looks like and you kind of have this template.

And you kind of have these specific patterns and what you are trying to do here is match this set template which is there in your head with the sensory input so you compare the incoming sensory input to a variety of templates you might have and you select a particular template that matches the sensory input to the best okay say for example you have templates A B and C and the input is let us say D you will kind of choose either of A B or C depending on the degree of match between D.

And either of these say for example a match a 65% the other two matches you know let us say 20 and 15% so you choose probably the you know a template of a because that matches D more than the other two okay in this template matching account we were looking we are basically looking for the exact match between the stored template and the input representation it is not even 65 or something the kind of thinking of a 100% match between this template.

(Refer Slide Time: 04:49)



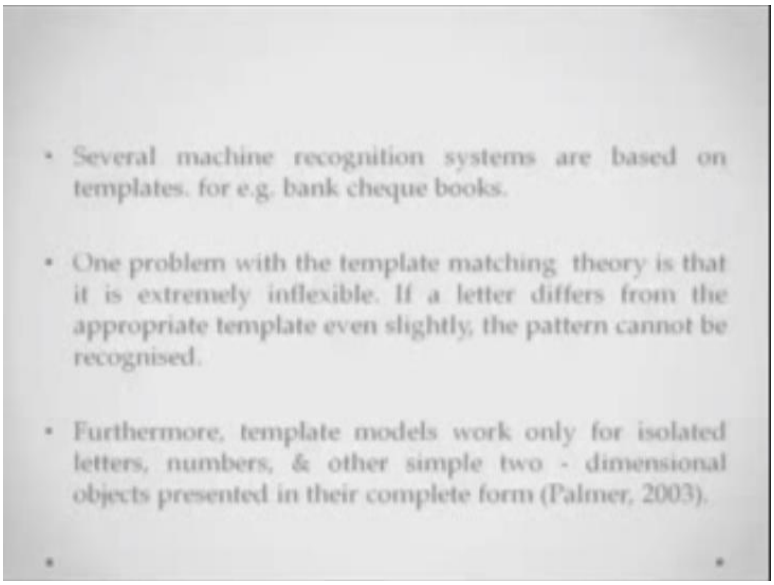
That we have stored and the representation or the sensory input that is coming in I will show you an example suppose you see you know here there are different ways of writing let us say pattern perception you will see that depending on different people is handwriting the templates of just

the set of alphabets can be really different now suppose you are a system which is supposed to recognize each of these things as soon as the template kind of varies by you know one degree or two degrees you will already start having a trouble in recognizing these patterns.

An example of these kind of things could be you know the, the Machine recognitions systems that examine for example your signatures in your you know check books if say for example you have just made a very small mistake in your signature you have kind of you know say for example you there is a superscript and there is a subscript if the subscript and superscript are kind of shifted a little bit the Machine recognition system which kind of matches.

This signature of yours with the signature that you have given at the time of opening the account and that matches less than a particular amount it will not recognize it okay so that is the problem with this template matching accounts the problem is that these are extremely inflexible theories if a letter were to differ you know from the appropriate template even slightly the pattern will never be recognized.

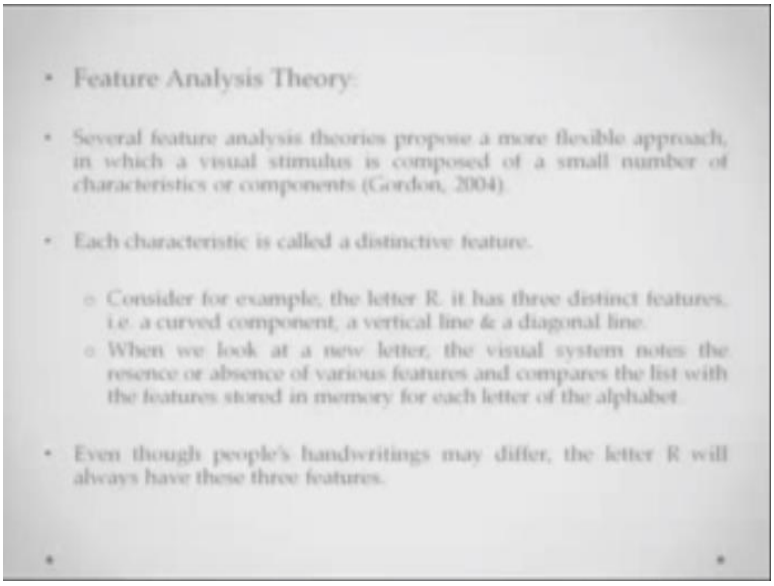
(Refer Slide Time: 06:04)



So but that degree of freedom is not there and you see that say for example our systems the you know recognition systems that we have are actually very good at it you know they do not really say for example if I am writing about delivered or you are reading a particular word and obviously given that our hidings will be really different you say able to read that word perfectly so it kind of tells us that maybe we are not really using a template and a template matching account and in that sense we let us we kind of move on to other models as well.

Now the template models in that sense by the way they are useful also but they are kind of you soon only and they work only for isolated net first numbers and other simple two-dimensional kind of shapes okay so if you have a really very, very simple kind of a setup obviously then may be the template matching account would work but not really for complex configuration say for example cursive handwriting for a large word that kind of brings us to a different kind of theory of object recognition this theory is known by the name of feature analysis theory.

(Refer Slide Time: 07:07)



Now this and this theory basically proposes that you know it kind of proposes a more flexible approach and this approach is about that any visual stimulus say for example you know a particular letter for that matter is supposed to be composed of a small number of characteristics

or component so it says let us not have a fixed template let us talk about the components that make up that particular object or that particular pattern and what we look for is not really the exact template but we look for the presence of these different components okay.

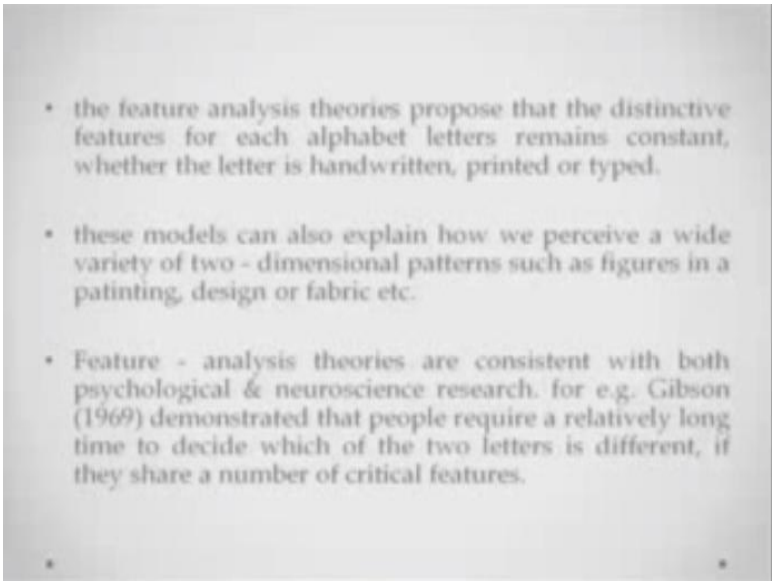
An example could be say for example if you talk about the letter R the letter R basically has three components in it has a vertical component it has a curved component and it has a slanted component now if I have a system that was supposed to recognize the letter R and irrespective of the you know the people whose handwriting I am kind of recognizing I would assume that the letter R we at least have these three features okay.

(Refer Slide Time: 08:31)

So in that sense I am not going to be troubled by people is handwriting because anybody who writes the letter R will at least write these three components okay so this is something which wherein you find that you know the feature analysis theory might be good he has this example you see so these are these different letters and this is basically how these features are ordinates you have straight horizontal vertical diagonal lines you have a closed curve you have intersections you have symmetry about them.

So these are basically kind of features which were shown by Gibson that you know how letters really differ from each other which is respect to these distinctive features and give some believe that this is how we are really you know recognize these later this is how even our a higher object recognition kind of you know mechanisms would work on.

(Refer Slide Time: 09:08)



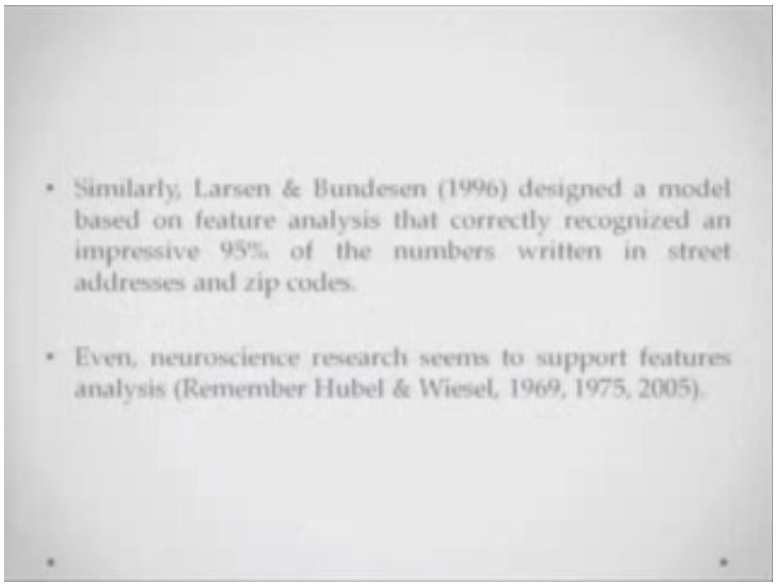
So the feature analysis series they proposed that the distinctive features for each alphabet remain constant irrespective of whether the letter is handwritten or typed or it is a photograph of a letter anything okay these models can also explain how we can perceive a wide variety of two-dimensional patterns such as figures in a painting design fabric those kind of things okay now a feature analysis theories as a group and there are many theories.

You are kind of discussing at a rather generic level feature analysis theories are consistent with both psychological and neuroscience research so Eleanor gifts we have been talking about our research and last things basically demonstrated that people do require a relatively longer time to decide if the two letters are same or different given.

That the two letters share a critical number of features say for example if you have decide between whether a particular letter is an N or whether it is an M you will see that these two letters share a lot of features they all share a slanted line and two vertical lines so the idea is because these two letters share these critical features you are matching of these letters or say for example the decision of telling that these two letters are different will take more time.

Because what you are doing according to this particular theory is checking for each feature so you check for air of vertical lines you find them alright you check for a slanted line you find them and find that as well but once you start looking for the second slanted line or let us say the second the different direction slanted line then you find that M has it but n does not and in that sense I am talking about a caps lock scenario.

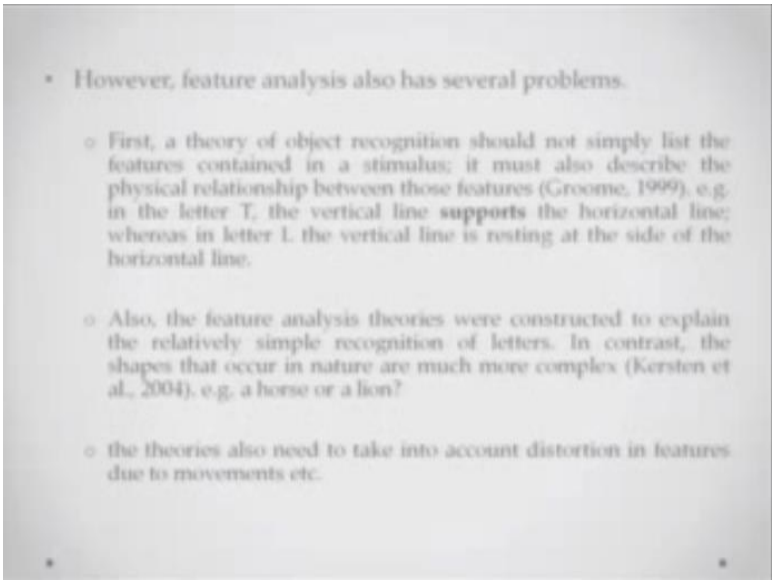
(Refer Slide Time: 10:54)



In that sense you will find that these two things are different this is what the feature analysis Theory says about recognizing these things so Larson in Bundoran they designed a model based on feature analysis that correctly recognized an impressive around 95% of the numbers written in street addresses and zip codes even neuroscience research also has you know shown that features you know feature analysis is something.

That we do you billion Wiesel did this research with neurons and they found that those neurons basically can be tuned to you know recognizing orientation say for example horizontal versus slanted versus vertical lines okay so you have these set of neurons which actually code for specific features in that sense you can say that feature analysis theory of object recognition has some support from the neuroscience data as well okay.

(Refer Slide Time: 11:41)



Now but feature analysis also have some related problems here there are some criticism on feature analysis theories itself let us talk about those problems now a theory of object recognition you know simply should not just list the features now the idea is if you are just talking at the level of features and isolated features you are not getting the entire you know how those features are linked with each other whatever in what way.

These features are jointed those kind of things you will not you are not really talking about okay also you can remember from the gestalt view of perception that you know an object is not really just components and components joined together just not really just give it one object it kind of the whole is more than the sum of its part thing okay so a theory of object recognition then

should not simply miss the features contained in a stimulus it must also describe the physical relationship between those features how are they linked together okay.

For example in the letter T the vertical line supports the horizontal line whereas in the letter L the vertical line is resting under the horizontal line as everything at the side of the horizontal line okay so you can look for a horizontal and a vertical line but you will find that it is present both in LnT so you need to really specify how this linkage is there only then you will be able to you know understand why say for example NL is different than the T now the feature analysis theories were constructed to explain the relatively simpler recognition of later see the object recognition in itself is a bigger problem feature analysis series.

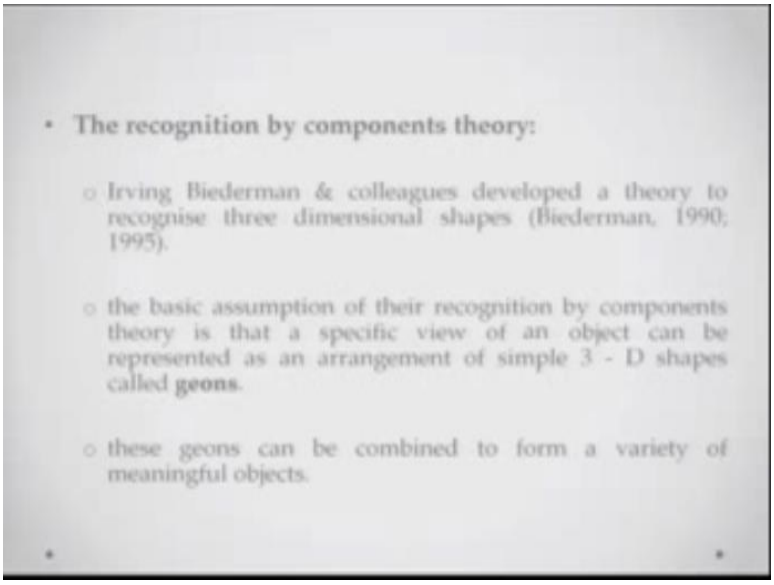
Basically started with the explaining us and the recognition of flatus so you know taking a very simple scenario there trying to solve a very simple problem but if you really look at you know objects in the world you know say for example plants and animals and automobiles and houses and those kind of things they are much more complex than just being a you know a concatenated set of features so in that sense a feature analysis.

You know kind of falls short of explaining the you know the myriad problems in understanding or recognizing different objects so theory also something that you see is that there are generally distortions you know things are moving and there are some kind of distortions in features as well say for example if you were to recognize what a cheetah looks like or say for example what the horse look looks like and you are looking at a horse while it is really running the features are changing you know and the percept or the sensory input.

That you are getting is also changing okay in that sense it will become that much more difficult for a feature analysis theory to explain you know what a particular object looks like so imagine say for example you know you have a particular presentation in which the features of the letters are mixing with each other or they are kind of moving say for example if you are reading something written on a flag.

For example you know the flag is moving with the wind and something that is actually horizontal does not look horizontal at the moment it you know looks slanted because the flag is removed by the air so those kinds of things the feature analysis theory we find slightly harder to explain okay so let us move towards a different kind of a theory another theory that we can talk about is having be demands recognition by component theory.

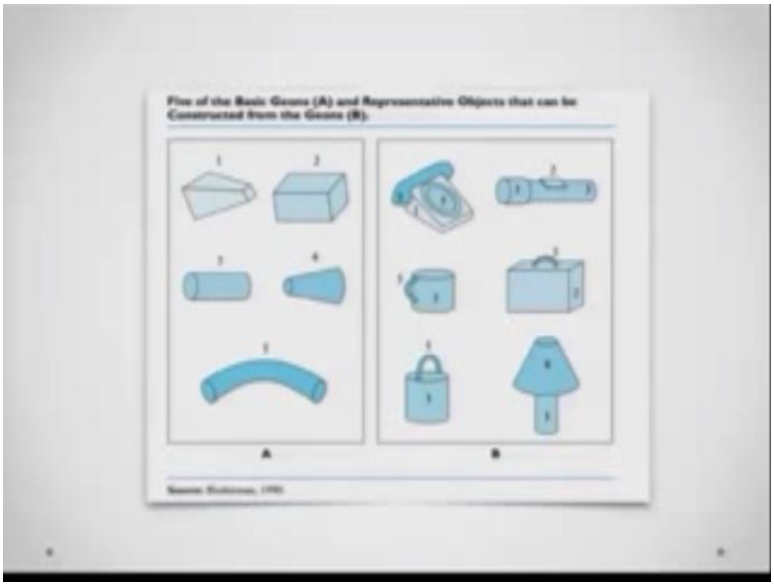
(Refer Slide Time: 15:09)



Now what is this recognition by companies theory B demand basically developed theory and to recognize three-dimensional shapes this idea was similar to what Gibson was saying we are not really dealing with two-dimensional entities in the external world we are dealing with three-dimensional entities so let us you know develop a theory that will explain understanding or recognition of three-dimensional objects rather than two-dimensional objects.

So with that the start the basic assumption of recognition by components is that a specific view of an object can be represented as an arrangement of a simple 3d objects and so Bide man basically came up with this simple 3dobjects okay Geons I will show you a Geons short while but these jobs were supposed to combine and represent particular shapes or particular objects and the idea was that we are actually understanding each of these Geons.

(Refer Slide Time: 16:08)

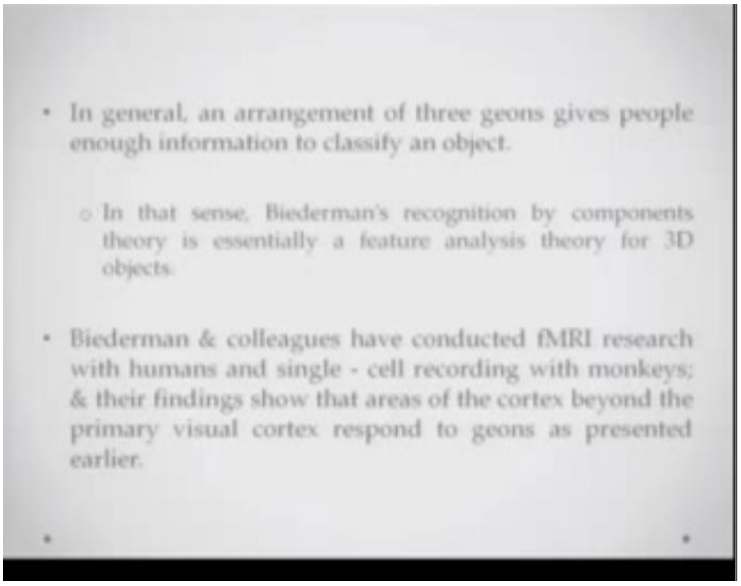


And that is what is leading you know to our understanding or recognition of the objects you can see here say for example you know you have Geons on the left side and you have these objects on the right side and you can see each of these objects on the right side are actually composed of these different Geons so you can see say for example the telephone is composed of Geons number 1 3 & 5 this thing the cup is basically made up of Geons number 3 and number 5 okay.

So this is basically you know in some sense you know you kind of extract this higher-level feature or say for example you can extract this higher-level component from the three-dimensional objects that you see and you make a sense of that okay if we combine this object with this object and say well you know the Geon number 3 and 0 number 5this will lead us to forming what is called a cup okay.

In some sense what you are doing is you are recognizing by these discrete components with just you know so all objects are basically you know can be understood to be permutations and combinations of these various Geons okay this is pretty much what you know the Irving be demand model of recognition by components was talking about in general let us elaborate on this a little bit.

(Refer Slide Time: 17:21)



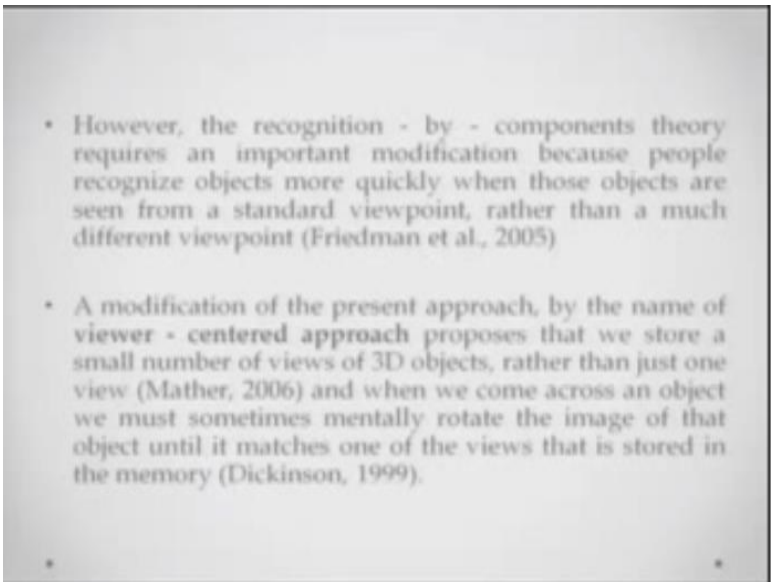
In general the arrangement of three Geons would give people enough information to classify a particular object in that sense Biederman's recognition by component theories essentially a feature analysis theory as I was already saying for three-dimensional objects. So you have three features you have a curb thing if you look at this figure again you have a cylinder which is number three you have a pipe like thing which is number five and you combine the cylinder with this pipe like things you get a cup which is an object so this is what is thing with this particular theory that you can combine these different components and these combinations are pretty much as features only.

And you can combine these different components to get representations of objects in the world so bitumen basically they conduct this research with humans and single-cell recordings with monkeys and their findings have shown that you know areas of the cortex beyond the primary visual cortex respond actually respond to Geons as presented earlier so the thing is we do have some coding for these three-dimensional features or three-dimensional components.

And in that sense maybe we are sensitive to how these components occur as parts of different objects and maybe that helps us recognize these different objects okay so there is some data

which supports the recognition by component theory now the recognition by component Theory also requires an important modification because people will recognize an object more quickly when those objects are seen from standard viewpoints.

(Refer Slide Time: 18:46)



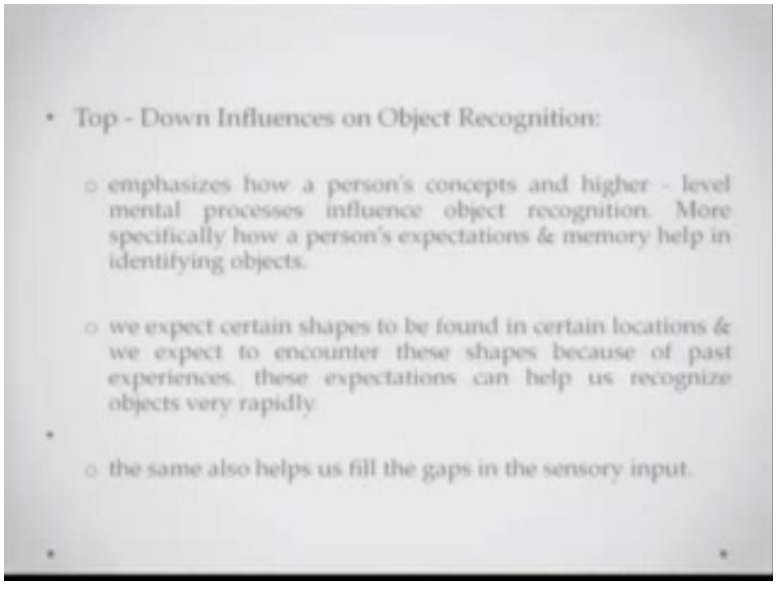
So now the thing is this if you see a cup or say for example if you see a phone here this is one representation of the phone or this is one representation of the car suppose I invert the cup suppose I kind of tilt the cup then the component shall analysis might change a little bit okay at least the feature extraction will change a little bit okay.

So you will need to modify this recognition by component theory by a little bit because people will recognize objects more quickly from a standard view point what is the standard view point it is a canonical view point okay rather than a much different or a non canonical view point now a modification of this approach basically named as the viewer centered approach this approach by the way recognition of components in the standard form is a recognition is the object centered approach.

But say for example if you are a person were moving around the room and having different views of these objects you would have to have what is called a viewer centered approach the viewer centered approach basically proposes that we store a small number of 3d objects as features or as components rather than just one view so we will just store it there you know a small number of views we will see what are the three or four views of a cup I can have you store each of these to your four views when you kind of develop a component shell analysis of how these three of you use will be made.

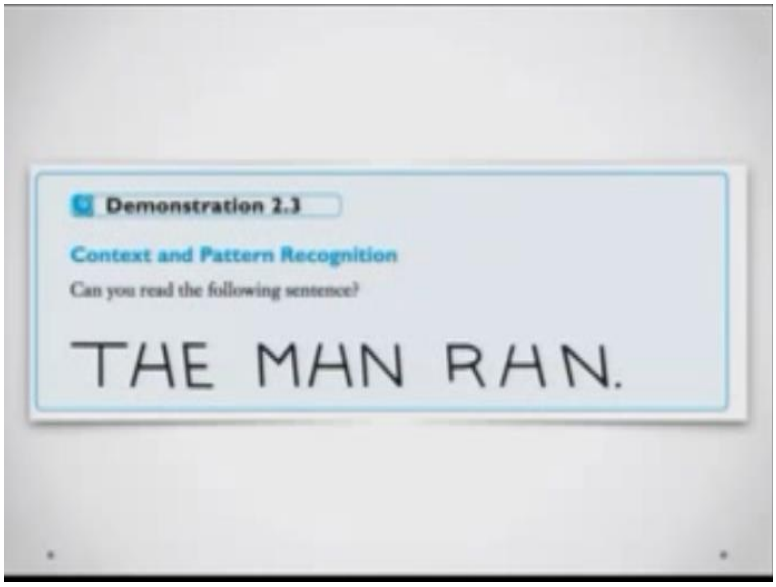
And that will lead to understanding these components okay so yes and when they when you come across such an object you mentally rotate the image of the object until it matches one of the views that is already stored and in the memory and then say for example by a combination of a top-down and a bottom-up Elche we recognize that particular object we cannot we talk about as you know it is a good point to talk about top-down influences now top-down influences emphasize how a person is concepts and higher level mental processes will influence object recognition.

(Refer Slide Time: 20:47)



More specifically how a person is expectations and memory may help you in recognizing these objects so we can expect certain shapes to be found in certain locations and you can expect to encounter those shapes because of past experiences say for example if you are looking you know at your study desk you will expect to find a particular say for example the notebook there or maybe a pen there suppose I bring a particular object and keep it may be I bring a cylinder you know and keep it if you touch that cylinder if there is no light you are not really looking at it if you touch this a little more often than not you will expect it to be a pen maybe it is maybe it is a pipe for that matter.

(Refer Slide Time: 21:32)



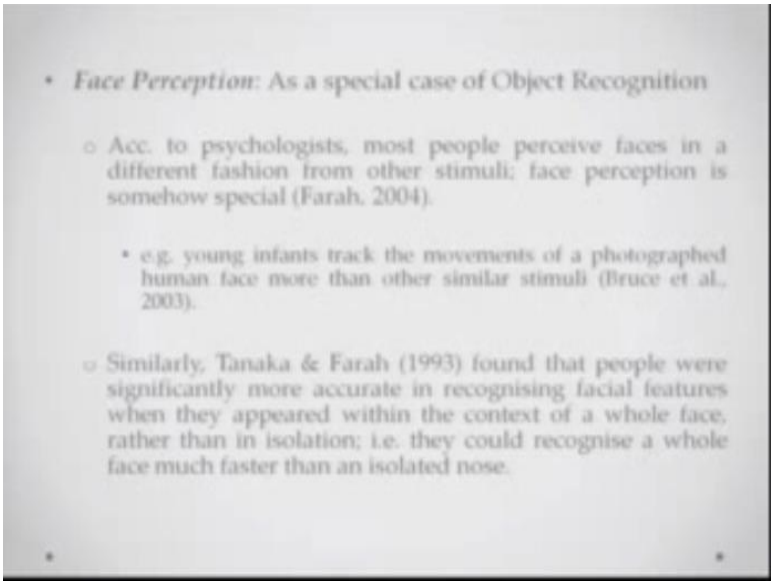
So these kind of expectations kind of you know help us fill the gaps in our sensory input an example could be something like this you see you can probably read this the man ran but you see both the A's have the top part cut so it is technically perceptually not an A but you will see say for example in the first case you read it as H in the second and the third case you know that this is a okay.

Even though it is exactly matching you know the H in the first word so you know what kind of thing to expect where and that is how the top-down influence will modulate your sensory experience and help you generate a perceptual representation something like this.

(Refer Slide Time: 22:12)

You know there is something figure which I message which I found floating on the Internet a lot of Facebook means about this and it says that you can read this now you can see say for example the first letter is basically replaced by a number but that number perceptually resembles the letter it was supposed to if you read this is the first word this is this the seven basically resembles the T and the five resembles the S.

(Refer Slide Time: 22:46)



So you can still understand it is this and say for example you talk about message the second word IRA's embolus s and the three resembles E and five resembles s so you can still make out that it is a message okay now let us move to a special case of object recognition let us talk about face recognition now according to psychologists most people perceive faces in a slightly different fashion than they perceive other stimuli and they say that face perception is somehow really special.

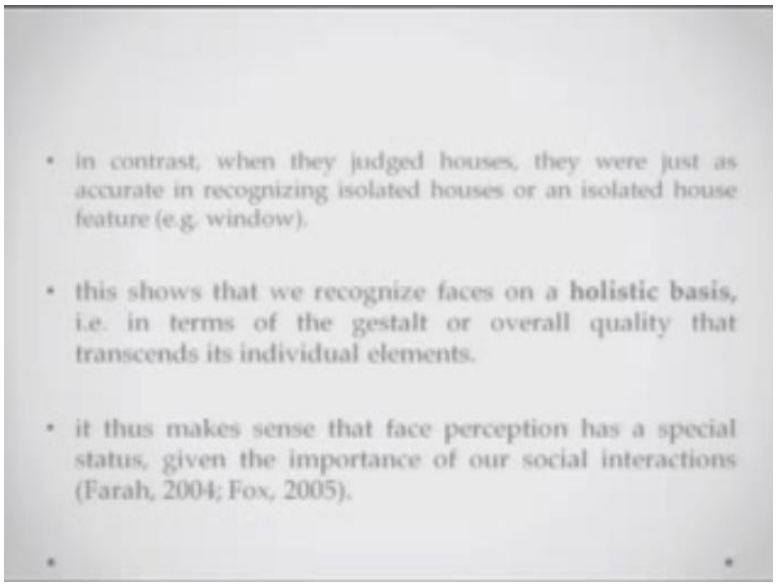
There is a lot of research into face perception and it all of it suggests that faces are slightly special stimuli as compared to other objects in the world young infants track the movements of a photograph human face much more than other similar simile say for example it shows that you know faces are socially or in some sense important to even the young infants you know you are evolutionally probably you know wired to treat the faces as a slightly special stimuli and it makes sense.

Because they say for example men being or humans being social animals this aspect of you know recognizing a face is a socially important skill also aspects of recognizing the emotions of a face is also socially important suppose you are looking at the face of an individual any suddenly

shoots you so how will that be understood if you're not recognizing the face of the individual that this individual is angry and you know might act aggressively if you are not really already on your feet you will you know you will pay a huge cost.

So in that sense you can understand that faces area socially important stimulus they are special types and a class of stimuli a lot of research really talks about that say for example Tanaka and Farah in 1993found that people were significantly more accurate and recognizing facial features when they appeared within the context of a whole face rather than in isolation so you have to really look at the whole face you are not tuned to recognizing features of faces without the whole faces again something that probably is not true.

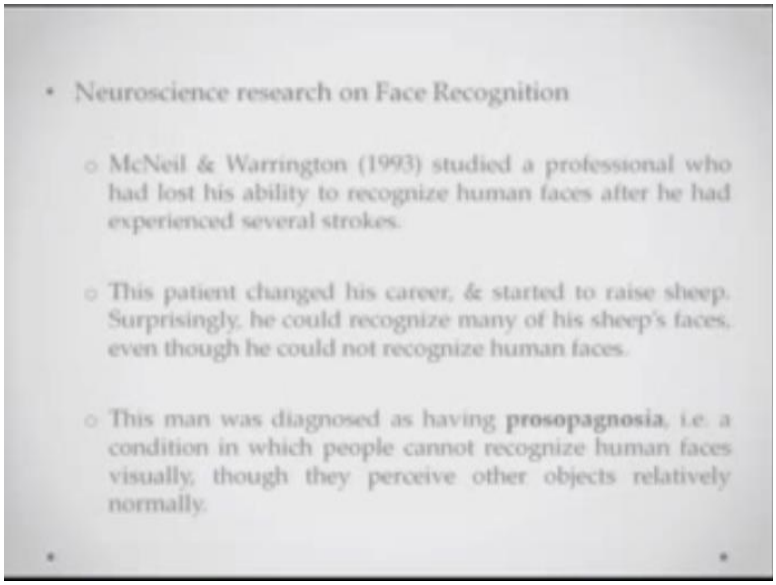
(Refer Slide Time: 24:43)



If you are talking about objects and features okay in contrast yeah that is the contrast I was talking about even the church houses they were just as accurate in recognizing isolated houses or an isolated house feature say for example a door or a window or a gate something like that this shows that we recognize faces on a holistic basis we kind of have a holistic or overall understanding of what a face should look like what are the features that holistically this particular object should have.

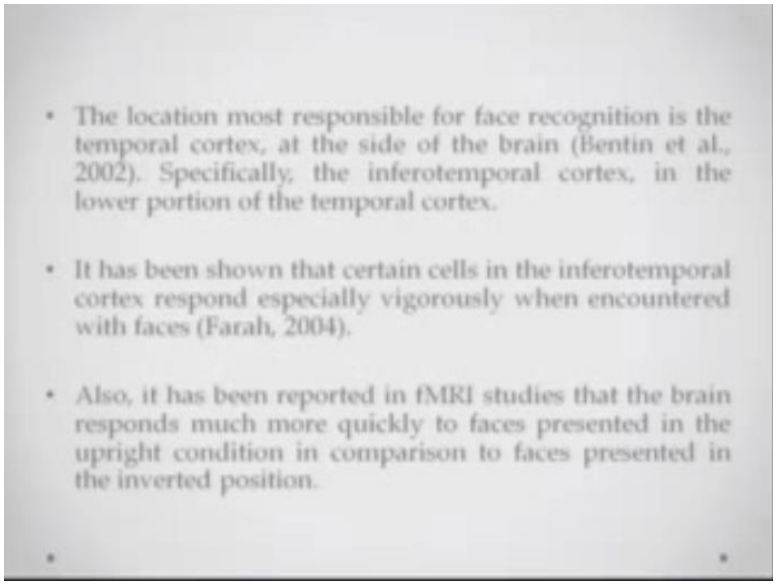
And that is how we kind of you know this organization of the eyes and the nose and the mouth and the ears is what we kind of look at as a face okay so it is kind of in that sense slightly difficult if the Isis are presented separately or if the noses in it separately it likes lightly more time for you to recognize that this makes sense because face perception has a special status given the importance of our social interactions something I was already talking about there's also a lot of neuroscience research on face recognition McNeill and Warrington.

(Refer Slide Time: 25:33)



They study the professional who had lost his ability to recognize the human faces after he experienced several strokes this patient at a later point changes career started to raise sheep but surprisingly it was found that he could recognize many of the sheep faces though he could still not recognize human faces now this special condition was diagnosed and later termed a sprosopagnosia that is a condition in which people cannot recognize human faces visually though they can perceive other objects relatively.

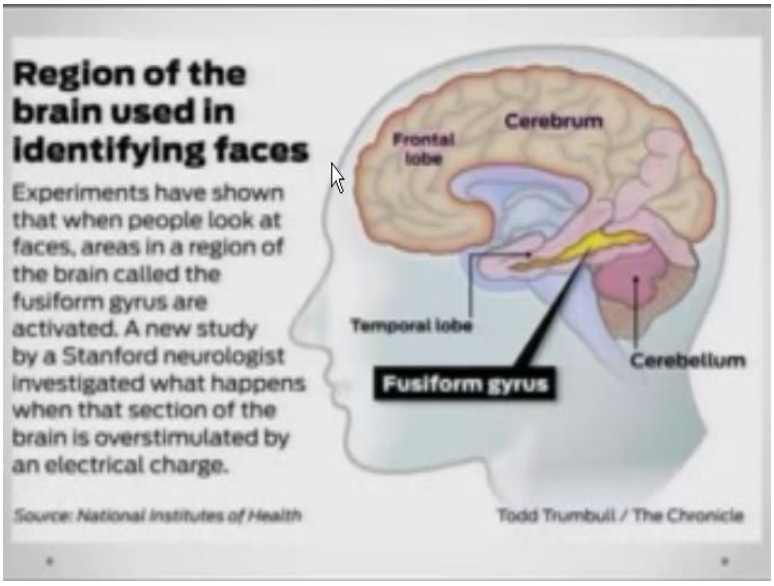
(Refer Slide Time: 26:14)



Normally so it again is a clue to how important recognition of faces is for us as humans the location in the brain most responsible for face recognition is the temporal cortex at the side of the brain generally the right side specifically the infer temporal cortex so the area under their temporal cortex in the lower portion that is you know what is implicated it has also been shown that certain cells in the infer temporal cortex response especially vigorously.

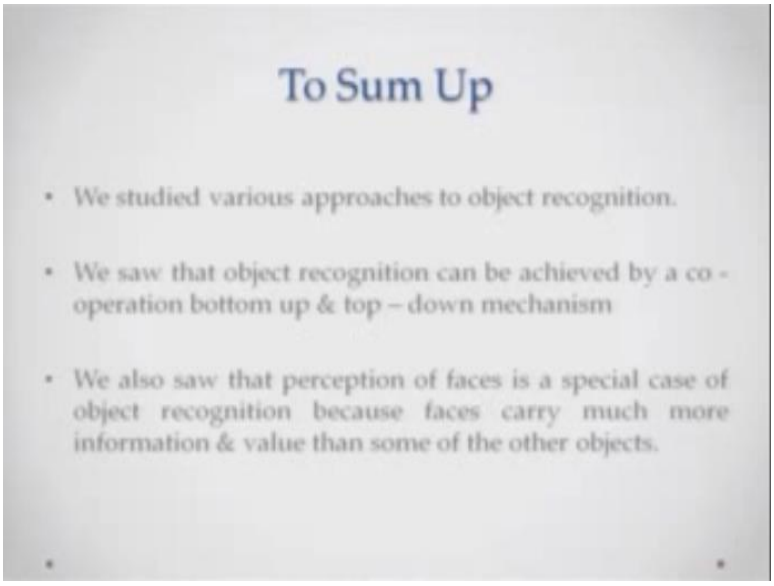
When encountered with faces so this is the ADA which kind of lights up when you are you know presented with a particular face also it has been report a lot of ephemera I said is that the brain responds much more quickly to faces present in the upright condition than to face is presented in the inverted position because the configuration would change actually so then you would probably have to apply a feature analysis or a recognition by components kind of approach to see what faces if it is automatically.

(Refer Slide Time: 27:14)



If it is in the canonical position you will probably gain more information out of it much more quickly so this is the representation this is the yellow region is basically called the prosopagnosia this is the region which is actually responsible for us to recognize faces now this was all about object recognition.

(Refer Slide Time: 27:24)



Let us try and sum up we have talked about object recognition we saw that object dimension can be achieved by a combination of top-down and bottom-up approaches we also saw that perception of faces is a special case of object recognition because faces carry much more information in social salience as compared to some of the other objects that we interact with thank you.

Acknowledgement

Ministry of Human Resources & Development

Prof. Satyaki Roy
Co – ordinator, NPTEL IIT Kanpur

NPTEL Team
Sanjay Pal
Ashish Singh
Badal Pradhan
Tapobrata Das

Ram Chandra

Dilip Tripathi

Manoj Shrivastava

Padam Shukla

Sanjay Mishra

Shubham Rawat

Shikha Gupta

K.K Mishra

Aradhana Singh

Sweta

Ashutosh Gairola

Dilip Katiyar

Sharwan

Hari Ram

Bhadra Rao

Puneet Kumar Bajpai

Lalty Dutta

Ajay Kanaujia

Shivendra Kumar Tiwari

an IIT Kanpur Production