Cognition and its Computation
Prof. Rajlakshmi Guha
Prof. Sharba Bandyopadhyay
Biotechnology and Bioengineering
Indian Institute of Technology, Kharagpur

Lecture - 34
Visual Search and Pattern Recognition

Welcome. So, we have discussed about object recognition and now we go on to the second phase of object recognition, in terms of the visual object recognition and that is Visual Search or Pattern Recognition by searching in a visual scene. So, we have said that attention plays a big role in object recognition or in general in perception as we have discussed. Because attentional signals ultimately modulate the representation, neural representations by contrast enhancement, that is by increasing the responses to a degree to of the attended region or attended feature.

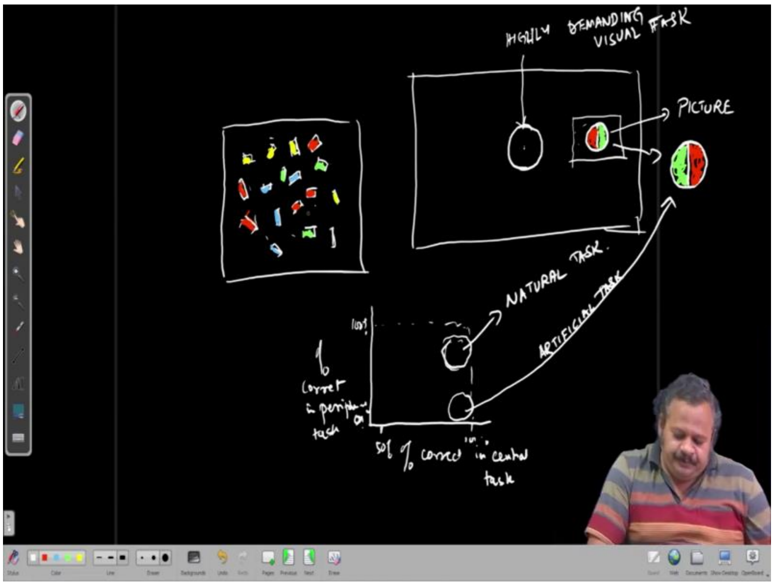
And reducing the responses of non-attended feature or the feature that is distracting and this kind of contrast enhancement lies beneath the attentional enhancement that is observed in performance when a human or a behaving animal is performing a particular task if it is engaged with attention as opposed to when it is not paying attention and the neural mechanisms behind that we have discussed in our attention lectures.

So, the visual search, the idea is that we have often neuroscientists have often studied visual search and attention in very artificial scenarios; however, visual search is highly important to study in natural scenarios. And in fact, we find that we humans or animals who are trained to do a particular task are extremely good and extremely efficient in performing object recognition tasks or searching for a particular object in a visual scene if it is a naturalistics in a scene that it is used to.

In other words, our system is actually biased to process information or identify and search and identify objects in a scene even auditory scenes, we are used to or biased to doing so in particular ways we have been doing it all along. That is the things that we have experience with and that sort of goes against the studies of, experimental studies of visual search and attention.

Although all the initial studies of attention have helped us fundamentally to understand mechanisms behind attention and so on, a very important aspect is removed from the whole study, if we are not studying it in a naturalistic scenario. So, in terms of visual search the idea is that, if we have to identify a particular object either we identify it based on how salient it is or based on our experience of the scene and. So, if we take a small example an artificial example, let us say we have a visual field here and let us say there are many bars that have particular orientations and so on.

(Refer Slide Time: 04:23)



And each has different color, not each some of them have different colors. Let us say it is something like this and if you are asked the question, that I mean if you were not seeing it beforehand and you were asked to identify, if there is a or rather search for a horizontal red bar. You would immediately identify it although there are many distractors, because it is salient and unique.

There is only one red among them. So, it is very easy to tell it apart and also horizontal, there are few horizontal bars there only 2 and so that makes it easy. The more we have the distractors sharing features, that are required for binding and identifying a particular object, the worse our performance gets. So, if the same scenario if we add bars here, that are also red in color of different orientations I am sorry of different orientation.

Let us say all the bars are of the same length and size and all that which is not true in my cartoon, but you will see that, it will as we as I increase the clutter here it becomes, it would become more and more difficult to identify a horizontal red bar. Now, that you know that which one is the red bar, but if you had not seen horizontal red bar, if you had not seen it earlier, I mean you can do this experiment yourself by asking someone else to look at the second image first and say whether they can identify or whether a horizontal red bar is present, they will take a longer time than when they are shown the other bar.

So, a visual search is heavily dependent on what distractors are there and how they are, how their features are related to the object that we want to identify. And the contrary to this idea is the fact that when we go into naturalistic visual search or into natural visual scenes, we actually do much faster than many artificial scenarios. So, we will start this idea with a specific example where an animal or actually this is with a human with humans.

So, a human is trained. So, if we consider a particular visual field human is trained. So, it in the central region in the person is required to focus in the central region and there is a highly demanding visual task that the person has to perform. Like some identification of letters in a jumble of letters, whether a particular letter is present or not, which the human does well once they pay attention to the central region.

If they are not paying attention to the central region they are unable to do that task ok. So, we have a highly demanding visual task at this center of the visual field which the human or the person or the subject has to perform continuously. And along with that in the periphery the person is shown scenes where the person has to say whether an animal is present or not, in a jungle scenario or so on. So, this is a picture whether there is an animal present or not.

So, what we find that a person actually when we look at the performance. So, let us say percentage correct, in central task and percentage correct in peripheral task. So, gradually, so this is 100 percent and let us say this is 100 percent and let us say 50 percent is chance. So, this is the minimum, this is the minimum. So, most people with many different kind of tasks they tend to be clustered around this region.

I mean not should not be greater than 100, they tend to cluster around this region, that is they are almost equally good at the central task and the peripheral task. When these same subjects are asked to do a peripheral task of where they are shown in the periphery, basically either a circle and that has a red color on one side and green on the other side or this or red color on the opposite side and right side is red green on the left side. So, the all the subjects are equally good at all the tasks when done independently, in the periphery and so on.

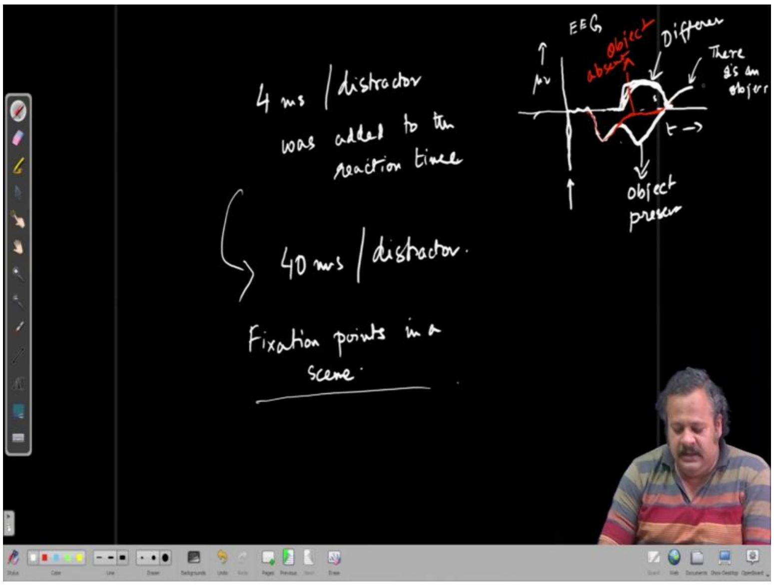
This task in the periphery with no central task, they do this 100 percent correct. When it is identifying whether an animal is present or not and when it is identifying whether the red is on the left side or the right side of the disc, the performance drops drastically. And so, in this case they are near chance. So, this is for the artificial task and this is a natural scene or natural task of identifying an animal in a jungle scene.

So, that shows us the power of the natural scene and how we are already wired to identify objects in particular contexts that we are used to. So, not only is it that the distractors are important, but also it is the context or the natural scenario the of that is behind the object recognition task that is also equally important.

In fact, when we go on to the natural visual scenes what so, this experiment is done in a setting, where a person has to identify a lamp in a living room scene. And humans are extremely good at identifying the lamp in a scene where there is that there is a visual, I mean there is a scene where it is the living room.

And it the performance does not drop down with increase in the number of, drop down drastically with increase in the number of distractors. So, in this particular kind of experiment, where careful controls were done in a naturals kind of scenario like the lamp in a living room, it was found that additional distractors on average different very different kinds of distractors were tried. Generally, about 4 milliseconds per distractor was added to the reaction time.

(Refer Slide Time: 14:34)



What we mean by reaction time is that the time the person takes to identify the lamp in the room let us say. Once the picture is shown to the person to the time the person identifies the lamp and provides a response, that is what the this duration is the reaction time. And it is found that in a natural scenario with many different kinds of distractors doing proper controls; we find that in the that kind of scenario only about 4 millisecond is added to the reaction time, which means the you are doing the task worse gradually worse and worse.

But by a very small amount 4 millisecond per distractors. On the contrary in general similar tasks with unnatural or artificial scenes, like the bars that I showed you earlier It the reaction time increase is 10 times, that is about 40 millisecond per distractor. So, this again shows us how we are wired or rather biased in some way to perform object recognition and or do the visual search to identify particular objects in different scenarios in environment that we are used to and that is that we are experienced with and.

So, following this we people have studied into what are the reasons behind it. And, so in order to look at that if you people have studied the fixation points in a scene. So, if you think about the bars example that we discussed in the previous slide. If one were to track the persons eye positions throughout the process of searching for the horizontal red bar, we would find that different people would be looking around at different locations initially to finally, pinpoint where the red bar is. So, if we look across subjects many

many subjects probably the entire visual field will have spots at which one or the other subject has looked at.

But when we look at natural scenes, let us say scene of a street where there is a bus going and maybe traffic light and some buildings on the side and the pavement and the seat on the pavement. And you are asked to look for a person, you will I mean there will be consistent looks at on the particular positions on that visual scene.

That is either on the pavement where the seat the seating area is or on the other side of the pavement, where people are walking or at in the crossing of because there is a traffic light. Here let us say there is a crossing people would be only looking at those positions for humans. Because we have a certain expectation of a particular of the particular queue that we are looking for or the particular object that we are looking for and we look for them in those particular locations and tend to find them.

So, as opposed to the artificial scenarios, where we actually search throughout the visual field or actually forced to search throughout the visual field and so, this gives us a very fast, this actually allows us to identify objects very fast. And so, that is also seen through EEG recordings, where when we compare event related potentials.

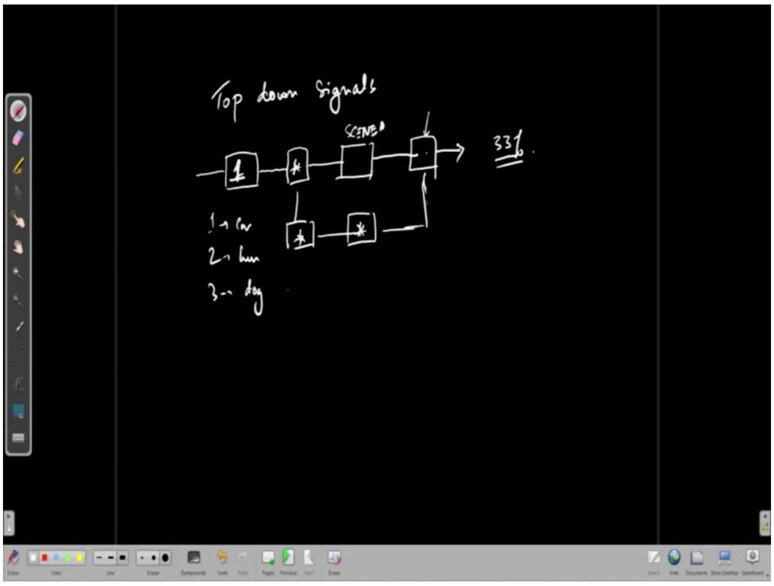
So, EEG signals in, let us say micro volts that we are recording from relevant electrodes and this is the time point ; so this is the time point when a particular scene is provided to the person and the person is asked to identify, let us say a person or an animal in the scene or something.

And so these event related potential, so this is baseline often look somewhat like this. And so, this is about the 200 millisecond time point. And so, this is the case where there is an object present. In the same scene, if the object were absent then we would get rather not like this, then the same event related potential would turn out to be, let me draw it in another color would turn out to be something like this.

So, in other words and this is about 150 milliseconds. So, the changes in event related potentials, so this is the difference that we see. So, the red more or less overlaps up to here and then here and this is the case where object is present and the red is the case of the EEG signal when the object absent. So, the person is shown a scene where the person is not provided the object that the person is looking for.

And the difference in the EEG shows that by 150 milliseconds we have already at a level where we have started to identify the object. So, this is extremely fast; however, this is much later when it is not a natural scene. So, are actually around 200 milliseconds so, around the 50 milliseconds gain in when the object is present. So, this shows that we are actually again the same point that we are wired already in some way for natural scenes to search for visual objects, for particular objects.

(Refer Slide Time: 22:32)



And so, the finally, there is another experiment that we need to talk about and that is or basically to look at the top down signals themselves. To actually show that this kind of expectation to find a particular object in a particular scenario is actually, because the top down signal that provides the that provides attention and hence enhancement of whatever we are looking for and guiding our visual search, that itself is different.

So, in these experiments where people were queued to say to find a particular object in a scene that they would be shown in a sequence, let us say there is first in the screen a number comes up, telling the person the kind of object to look for. Let us say it is a car or a human, each number would mean something or let us say 1 is a car, 2 is human, 3 is a dog and so on.

And so the person knows that when one comes that person has to find a car and then basically there is a queue, there is a blank queue and then finally, the scene is shown and the person has to react and find and actually say whether the object is present or not and do it correctly.

So, this is done in 33 percent of the trials. So, out of 100 times a person is shown one of these numbers and is asked to identify and is asked to search for a car or a human or dog, in the scene here only 33 percent of the time actually an object is present. In other cases there is nothing that is shown, blank and during this period the signals through FMRI are captured in from the prefrontal cortex.

And what is found is that given the scenario, given the scenes or and given the q the reaction times that were obtained are heavily negatively correlated with the degree of similarity of the scene. And the prefrontal top down signal and that basically shows that the top down signal itself is priming the bottom regions that is the visual regions to look for a particular thing in a particular scenario.

Because it knows that this is what is supposed to come in a scene like a car is supposed to be present in a particular kind of scene. So, this kind of an experiment allowed one to differentiate what the top down part of the signal is doing in such visual searches. So, in other words what we conclude here is that, it is not just the bottom up visual representation and object recognition circuits that are involved.

But it is also the frontal cortex or the rather the prefrontal cortex signals, how they prepare the bottom regions that is from v 1 up to the IT regions that is also equally important in visual search and identifying objects. So, with this we will conclude our discussions on visual object recognition and visual search and pattern recognition. We will have a discussion on auditory scene analysis in the next lecture which is akin to object recognition in the visual domain.

Thank you.