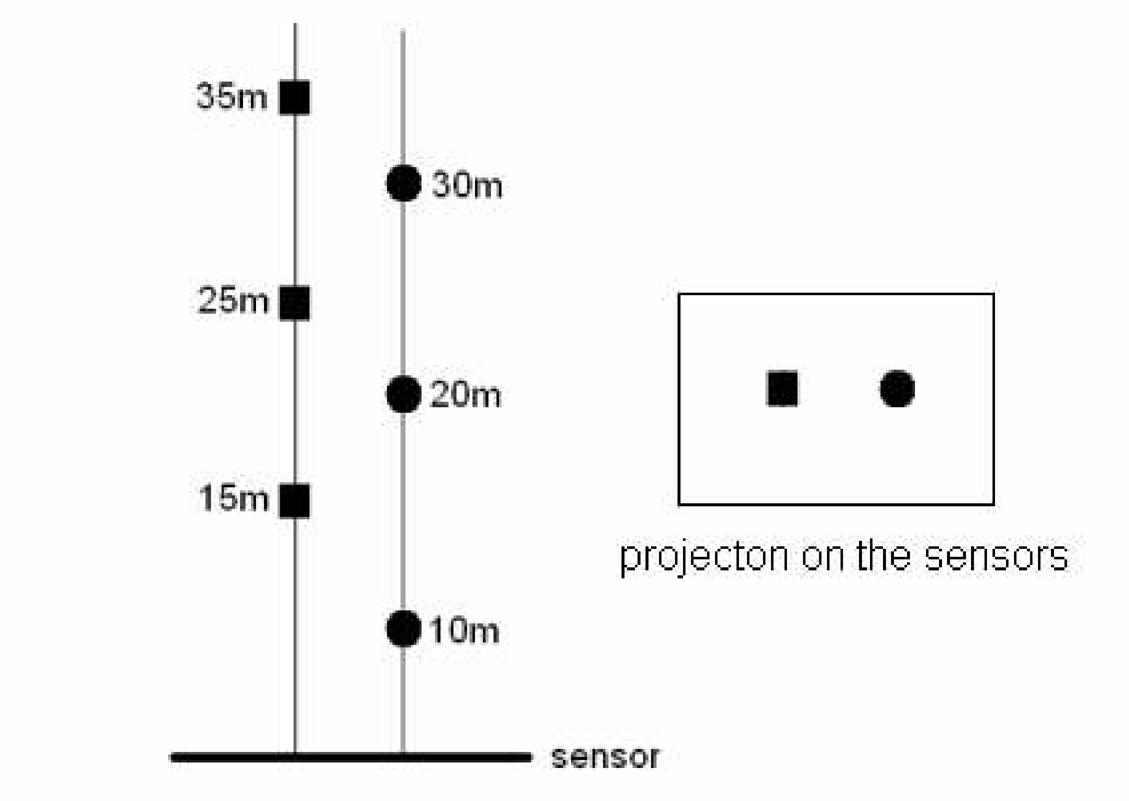
DEPTH PERCEPTION THROUGH STEREO IMAGING

Computer Vision (1):

I have been doing research in the field of vision since the last few years but a universal solution is nowhere to be seen; not only in my pockets but in any of the research institutes as well. Sometimes I feel it is impossible to find one universal solution to solving depth through stereo. Researchers have been thinking very deeply about just the stereo correspondence problem from many decades, but in vain. Either the problem is very difficult or we have missed the right track at the very beginning. I believe in either of these. The problem might be very difficult because it has taken its present shape through evolution, over millions of years. When I say evolution I refer to Darwin's theory of "The survival of the fittest". All these complex biological systems have taken rigorous stress test from nature and have survived to date, and to crack it might be a very challenging task. On the other side of the court we have man who has been able to build high speed miniature components and complex systems which should easily be able to replicate these relatively low speed systems. This is because biological vision is seen in small organisms like the insects, which hardly have say a few thousands of neurons dedicated for visual processing. Aren't our GHz processors able to achieve what has evolved in these small creatures? After so much of research and thought at least I believe that it is impossible to achieve depth perception through the currently tried out image processing techniques. Towards the end of my discussion I will try to take you on a walk along what path I believe can solve this problem. It might not be practical as of today, but who knows what technology is waiting for us at the door. If the current image representation and processing techniques are so poor at achieving recognition, why are all those researchers glued to it even now? That's because we always take our brain as reference for developing any recognition system and our brain is still able to perceive depth given a 2D stereo image pair. I believe the reader understands what a stereo image pair is. If not Google it right now! Or just skip it for now; I am going to take you all on a long journey covering each and every topic related to stereo images, depth perception, etc, etc. A lot of questions and answers were brought about during my research period and I have tried to give my best possible solutions to all of them. The whole idea behind writing these articles is to share these Q&A's so that a person fascinated about vision today will be able to start off from a much thoughtful point rather than repeating the experiments again and again. To race against nature we have to make sure we compress those millions of years of evolution into a much smaller duration.

Computer Vision (2):

First of all why are we so fascinated about our ability to perceive depth, or for a layman what does all this mean? After having vision (eyes) for so many years imagine a world without it. Frightening, right? Imagine having sight in just one eye. Most of them will be okay with it and some even ask me, what difference does it make? Now this is really frightening to us; computer vision researchers. We have been chasing this problem since so many decades, many researchers have even spent their entire life in vain trying to decode it and here we have some people who do not know its significance in spite of using it. No problem, what's this article for, then? There are two major things involved in vision; sight and depth. Many of them fail to distinguish between the two. Sight is the perception of light, and depth is the perception of the space around you. "An experience is worth reading 1000 pages", so better try it out yourself. Right from the time you get up in the morning spend the entire day closing one of your eyes. Observe if you can live life as easily as you could with two eyes open. (Disclaimer: I own no responsibilities for any accidents that might happen as a result of performing this experiment). But to get a feel of what is driving so many people in pouring so much effort for giving a machine the perception of depth, you got to try it out. Do not read my other posts till you have got at least something from this activity. One experiment that I don't want you to miss out is here: Hang a rope, a wire, stick, anything from a point such that there is space all around it. Get your fingers ready in the wire grasping position and move your hand towards the wire in the direction perpendicular to it to and grasp it. Remember to close one eye! If you get it right believe me, you are the luckiest person. If not, you would definitely want to know the magic that your brain is doing with two images. That is exactly what all our research concentrates on. Also try judging the depth between two objects placed at different depths with just one eye open. Try experimenting on as many objects as possible. It is impossible for you to know the distance between two objects without opening two eyes, except from monocular cues (I will come to this later). If you think about it carefully, there is nothing new I am talking of. When I say one eye, it is equivalent to taking an image from a camera. In a camera image the 3D surrounding is projected on to a 2D surface. From just this projection it is impossible to know at what depth the object was originally. Take a look at the image below. Square and circle are two objects in front of the sensor. Assume they are initially placed at (circle) 10m and (square) 15m. Their projection on the sensor would be as shown at the right. Try placing the circle anywhere along its line and also the square along its straight line. Do you see any difference in the place where they are projected? Not at all, you get the same image irrespective of where the two are relative to each other along their respective lines. Some people argue with me saying you should definitely be able to observe the change in the size of the object on the sensor as it moves far away from the sensor, so in some way you know whether the object is far or near. I totally agree, but what difference does it make? Who knows the size of the objects? I just have its projection with me at a particular instance of time and nothing else. When I move the object closer to the sensor, the size of the object definitely increases, but here we are talking about depth between two objects, which our brain accomplishes with two images. Even if the size of the object changes as you move it away or towards the sensor how does it give you the absolute depth of the object? We can always solve for two distances and sizes of objects such that one is big and far from the sensor and the other small and closer to it, both giving the same projection. Looking at the sensor you never know where the objects were because you don't know their actual sizes!

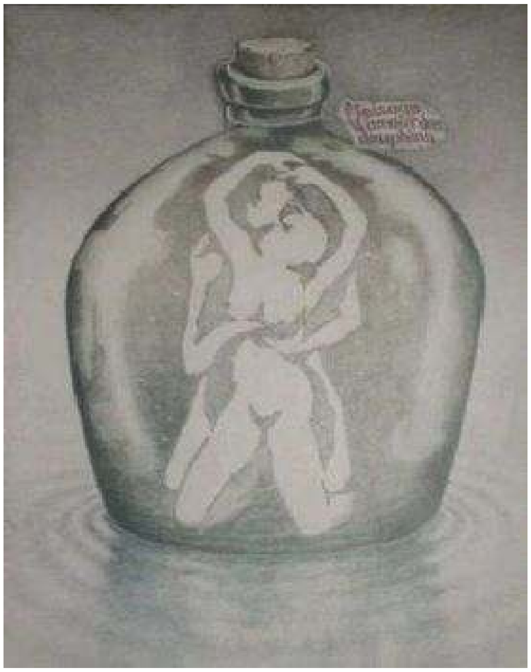


When you look at a photograph you almost get to know the depth associated with it due to a lot of monocular cues that your brain uses along with the knowledge gained over the years. I will have a separate post on monocular cues, so wait for that.

Computer Vision (3):

One statement that I usually get to hear from people is: "I understand the significance of depth on our perception of the surrounding, but a photograph as you said is a 2D image and my brain still manages to extract all the information from it. So should robotic vision depend so much on depth? Why can't we do away with it? Also take movies for example, which are a sequence of 2D images. You can actually feel depth in them, don't you? I still don't understand why we need 3D?"

We understand a 2D image completely because of our previous knowledge and not necessarily due to the image processing happening in our brain. Previous knowledge does not mean that we should have seen the exact image before; it means that we are aware of the context and content. Like in image processing, we don't understand the image after segmentation; both of them go hand in hand in our brain. What kind of segmentation our brain uses is still not very clear, but I can demonstrate how knowledge rules over the kind of segmentation that we can perform on an image. Look at the image below for example. What do you think the image contains? I am sure, 100% of the people would say "a man and a woman". You are totally wrong; the artist had actually drawn jumping dolphins on a pot! Now that you know the content of the image (knowledge) you can easily extract the dolphins out.



http://www.curiouser.co.uk/illusions/opticalillusions/dolphin.htm

You feel the perception of 3D in a cinema due to motion; a cinema is a motion picture! Motion can be obtained in two ways; one by keeping the camera static and having motion in the subject or bring about motion in the camera itself, irrespective of the subject. What our brain does using two eyes could have been done with a single eye, by oscillating it left and right to get the two images that it needs. The only difference would be that the images would not be from the same instance of time. From the time we start learning about our surroundings it is 3D vision that helps us segment the objects around us and put it in our database. Once we have gained sufficient knowledge about our surrounding we do not need 3D to perceive them, which is why we understand a 2D photograph without any problems.

I will be dealing with these topics in detail later on under illusion and 3D perception through motion. I have only been introducing you, to all of them now.

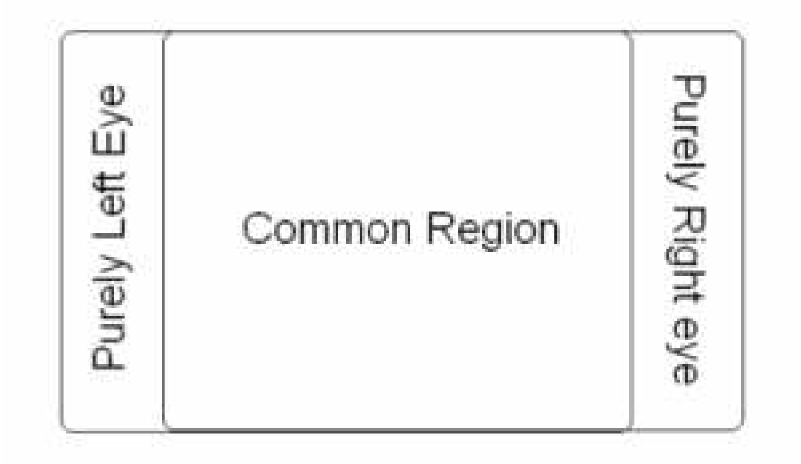
Computer Vision (4):

We humans have 5 different kinds of senses; touch, smell, sight, hearing and taste (correct me if I missed out something). We have one tongue, two eyes, two ears and two nostrils and of course skin for the sense of touch (skin is a special case I will come to it later). Ever wondered why we don't have two tongues? Does this number two or one make any sense to our senses? Let me illustrate their significance with some examples.

1. You can pick up a pen that is lying in front of you at one go. (Vision)
2. When someone calls you from your left you immediately turn towards your left instead of searching for the voice all around you. (perception of sound)
3. And of course fragrance definitely attracts you towards it. (sense of smell)

Each of these senses is highly developed in the order mentioned. As you can observe in these examples, when you have a pair of sensors they answer the question WHERE? WHERE is the object, WHERE is the sound and WHERE the smell is coming from? You don't have two tongues because you know that to taste something you have to place it on your tongue and can't do it wirelessly. WHERE, is something that becomes obvious in this case. The final sense is touch and when it comes to skin there is nothing like one and two and it covers our entire body. But we all know that it is sufficient to touch us at one place to feel it, rater than at two. You have to make a contact to have a sense of touch which eliminates the need to answer the question WHERE?

Just the presence of two senses needn't always guarantee the answer WHERE, it is their placement that gives it an extra edge. In general there needs to be some common signal that passes through both of the same kind of sense. Light is a high frequency wave and cannot bend along the corners. I mean you can't light up your room and go outside behind the room wall to read something, while this is not the case with sound or smell. So irrespective of where on your head the two ears or nostrils are placed, common signals will definitely reach them but you can't place one eye at the front and one behind your face. Light can't bend so you don't get any overlap or in other words a common signal in both the eyes. We humans have both our eyes on the front of our face, so it's very easy to get common signals. Want to experiment? Fix the position of your face and close one of your eyes, say the left one first. Remember the region that your right eye is seeing. Now close your right eye and open the left and compare the two regions. Most of the region that one of your eyes sees will also be seen by the other, which is the common region. The right eye will not be able to see the left most portion of the region seen by the left eye and vice versa. It is in this common region that we perceive depth. How? I will explain it in detail later, for now you just need to remember that "2 sensors == 3D perception".



Computer Vision (5):

If you closely observe the different species in the animal kingdom you will see that there are two kinds of creatures. Ones, that do perceive depth through vision and others that don't. The ones that don't will have their eyes towards the sides. Haven't seen such a creature? Give a deep thought; you would have even painted them in your childhood drawing classes. Fishes have their eyes towards the sides and hence cannot perceive depth. Then how do they move about, you were finding it very difficult with just one eye open? Won't their survival get affected from it? Not really! Instead it is evolution that has given them such an eye sight just for survival.

In general the observation is this; predators have their eyes towards their front and prey will have them on their sides. Let's take the ferocious tiger for instance. A tiger needs to pounce on its prey exactly and can't use the trial and error method that you used to catch hold of the wire :) (Refer my earlier post). To take a decision it needs to know the exact location of its prey which is given by depth. For a prey on the other hand, it is enough to know the presence of the predator; its exact location takes a low priority. For the predator its focus is on the prey and not the surrounding, for a prey its focus is on the surrounding, because it needs to look out for any possible danger from all sides. Evolution has hence given a predator a narrow angle vision but an overlapping one to perceive depth, while the prey has a much wider angled vision but lacks depth.

This does not mean that a prey does not have depth perception at all, it is just that wide angle is more important than depth. So the overlapping region is very small between the two eyes. Its face is designed in that way. We for example, along with say a tiger and other predators have a flatter face to hold both the eyes on the front, while a deer for example has a curved face so that their eyes are somewhat towards the sides. There is also a special case, creatures that play a dual role of predator and prey. Chameleons have adjustable eye sockets. When they sense danger the sockets move towards the sides to get a wider look, and while hunting they come closer to get an overlapped view! In the overlapped view both the eyes look at the same object while in the independent view they can process the two images separately. In our case that is not possible. Even though we have two eyes we cannot see two different objects at the same time, our eyes cannot move independent of each other.

Computer Vision (6):

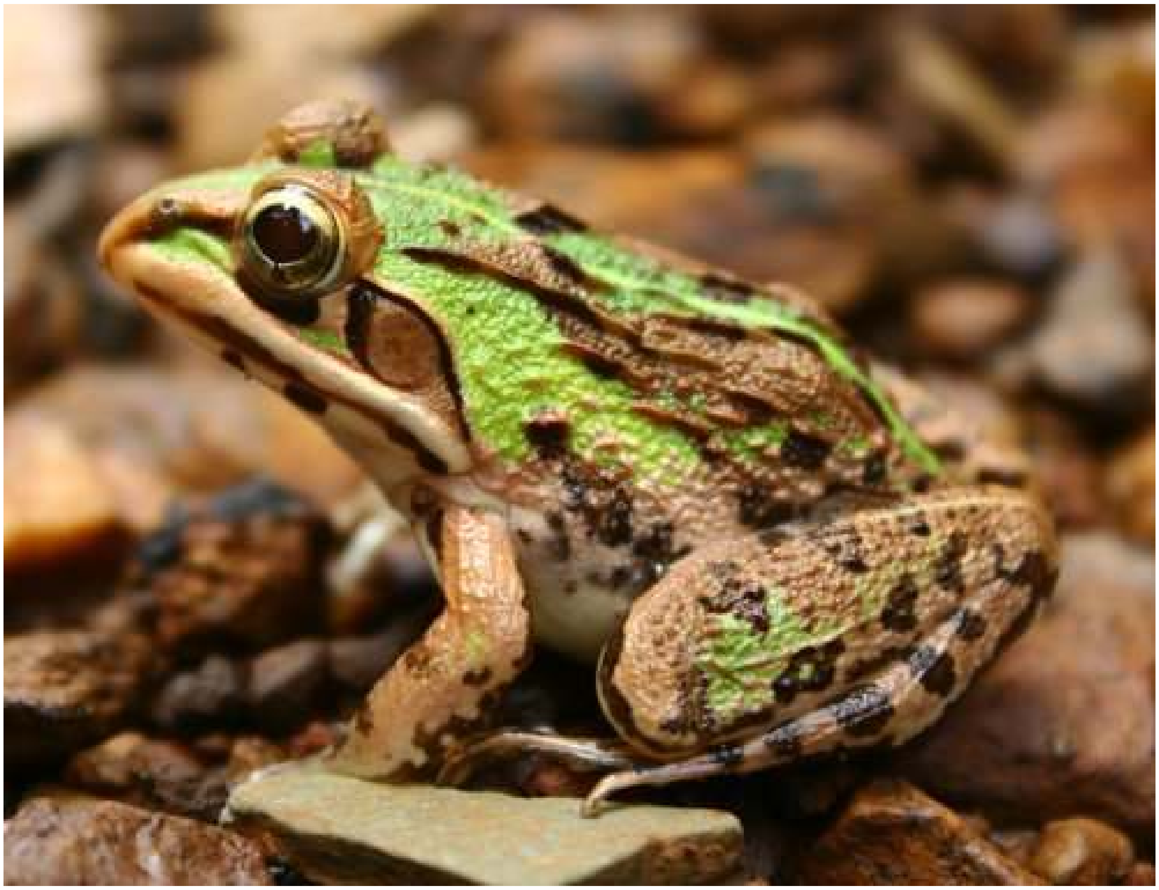
In my earlier post I was saying that even though we have two eyes we cannot use them independently. If our eyes cannot move independent of each other, what is it that is holding them? For both the eyes to see the same object either our brain has to be doing some kind of correlation between the images and providing feedback to the eyes asking them to position on a common point or the eyes themselves know where they have to be pointing. I mean, either it is a process of learning or it comes along when our system (life) is booted. This is actually a debatable topic. I tried to find an answer to this by observing it in small babies, but haven't been successful enough to conclude. Anyway I have some other observations to share. Depth perception does not produce an interrupt in the brain like the way sound, motion or color do. During the initial learning stages it is interrupt that matters because you need to draw the attention of a baby's brain to observe something, so depth takes a back seat. I term it is an interrupt because it immediately brings your brain into action. In order to achieve this you generally tend to get some colorful toys that make interesting sounds and wafture in front of a baby. So how does it work?

Sound, as you know definitely produces interrupt in your brain, which is why you use an alarm to wake up in the morning. Colorful objects produce high contrast images in your brain which are like step and impulse functions; strong signals that your brain becomes interested in. Now you know what kind of dress to wear to draw the attention of everyone around you!

If you remember awakening a day dreamer by wavering you hand in front of him, you know how motion produces interrupt in your brain. This is actually because of the way our visual processor and retina are designed, which I will come to shortly. So next time you are buying a toy think about these. Secondly why interrupt matters is because the new born baby's brain is like a formatted hard disk, ready to accept data, but has nothing. When it doesn't understand anything around it, there is absolutely no meaning in perceiving depth. Whether it perceives or not, it is just going to be a colored patch and nothing else. Again it wouldn't know which color it is! So interrupts help it to make sense of its surrounding, and when that is done depth and motion help it to segment the objects from one another to form its database.

Computer Vision (7):

This is with regard to the post where I was talking about the design of the visual system for predators and prey. I had taken an example of a chameleon for my explanation but could not photograph it on time :(. A frog never the less is an equally good specimen for this dual role played by many creatures in nature. Moreover, I had got a really good side shot of a frog when I had been to Agumbe and wanted a reason to share it. At that time I didn't know I would be coming up with this blog, or would have photographed the FV also. I thank **Kalyan Varma** for allowing me to use the FV of the frog captured by him here.



http://www.flickr.com/photo_zoom.gne?id=427866352&size=l

Observe the design of its face and eyes. The eyes are placed at almost 45 deg to the face. In the side view it can keep an eye on almost one complete side of its body and watch out for predators.

http://kalyanvarma.net/photography/photo.php?id=56&tag=Amphibians

In the front view you can see that they still have some room for common visual region in order to perceive depth which will be used to strike the prey with their tongue.

Computer Vision (8):

The reader, by this time would have understood the problem at hand and also how we are looking forward to solve it. If not, just keep a few things in mind. With just one eye, it is not possible to perceive depth. Without depth, it is not possible to segment the objects around us so effectively. Without segmentation it is not possible for us to learn or update our knowledge. If you have got the essence of it, some of the questions that would definitely pop in your mind are:

1. Is solving this problem so difficult?
2. Why would we want to solve it the way our brain does, isn't there a better way?
3. When a camera auto focus system can estimate the depth using IR, why can't we use say LASER to get the exact depth?

To explain why we would want to solve it in the same way as our brain does, I would like to quote these lines taken from the introduction section of one of the related papers from MIT. It states, "The challenge of interacting with humans constrains how our robots appear physically, how they move, how they perceive the world, and how their behaviors are organized. We want it to interact with us as easily as any another human. We want it to do things that are assigned to it with a minimum of our interaction. In other words we can never predict how it is going to react to a stimulus and what decision it is going to take.

For robots and humans to interact meaningfully, it is important that they understand each other enough to be able to shape each other’s behavior. This has several implications. One of the most basic is that robots and humans should have at least some overlapping perceptual abilities. Otherwise, they can have little idea of what the other is sensing and responding to. Vision is one important sensory modality for human interaction, and the one in focus here. We have to endow our robots with visual perception that is human-like in its physical implementation. Similarity of perception requires more than similarity of sensors. Not all sensed stimuli are equally behaviorally relevant. It is important that both human and robot find the same types of stimuli salient in similar conditions. Our robots have a set of perceptual biases based on the human pre-attentive visual system. Computational steps are applied much more selectively, so that behaviorally relevant parts of the visual field can be processed in greater detail.”

I think that this completely justifies the claim made above. For us what is important is how useful it will be for us humans. Take for example, the compression algorithms used in audio and image processing. Audio compression is based on our ability to perceive or reject certain frequencies and intensities. It is compressed such that there won’t be any perceptual difference between the original and compressed data for our system. For a dog it might really play out weird! Image compression also works on the same basic concepts.

As you go on reading my posts you will get to know whether the problem is difficult or not (that is the main reason why I started writing). I can’t answer this question in one or two lines here. Coming to the third question, LASER will always give you an exact depth or distance of an object, but our brain doesn’t work on exactness. Even though your brain perceives depth it doesn’t really measure it. Secondly getting intelligence out of a LASER based system is a tough one. If you use a single ray to measure the depth of your surrounding, what if your LASER is always pointing on an object moving in unison with the LASER? We need a kind of parallel processing mechanism here, like the one that we get from an image. The entire surrounding is captured at one shot and analyzed, which a LASER fails to do. You cannot use multiple LASERs, because in that case, how would you distinguish the received signals from the ones that left out. The ray that leaves the transmitter a particular point need not comeback to the same point (due to deflections). In that case what will be resolution of the transmitters and receivers or how densely should we pack them? What if there was something we wanted to perceive in between this left out space? This is neither the best way to design a recognition system nor a competitor to our brain, so let’s just throw it away.

Assuming that evolution has designed the best system for us, which has been tried and tested continuously for millions of years, we don’t want to think of something else. We have a working model in front of us, so why not replicate it? And this is not something new for us; we have designed planes based on birds, boats based on marine animals, robots based on us and other creatures, etc, etc.

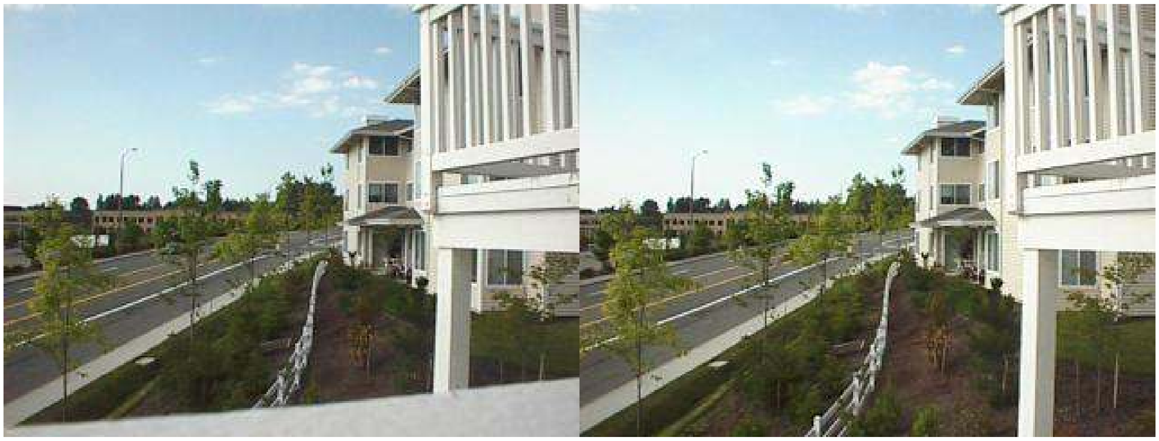
Computer Vision (9):

For a temporary moment let's forget about 3D and depth and concentrate on one more observation related to the angle of view and the kind of image sensor we have in our eyes. When you open both your eyes you get nearly 180 deg view of your surrounding, but how much of that 180 deg can you really see or perceive? Not much, here's how it is. There are two kinds of specialized sensors in our retina; one that is necessary for perception and the other specialized for detecting changes. The sensors responsible for perception are placed at the center of the retina in a region called the **macula**. These sensors are densely packed in this region and are responsible for clear vision necessary for reading and perception. To prove this, look at a particular word somewhere at the center of the page and try reading the line at the top of the page. Even though the entire page falls in the region of the visual field, you can't really get a clear picture of whatever falls outside the macula. Cats don't have this region at all and therefore how much ever you try, you can't train it to read something. In the remaining region of the retina lie the other kind of sensors, which are responsible for detecting motion or changes in the surrounding. When we wave our hand in front of people to get their attention (when they are lost in some deep thought) it is this region that interrupts their brain. In our day today life we fail to observe these minor things and feel as though we can clearly see the entire 180 deg around us. So if at all you have to see something clearly you have to position your macula over it. That is the reason our eyeballs keep swaying as we look at different objects. Observe this right now!

Computer Vision (10):

Let's come back to depth, which is our main topic. To perceive depth we know that both our eyes need to capture some common region. Even in this common region you cannot see two different objects at once even though you have two eyes. Try it out right now! Take up a long word and try to see the first and the last characters at the same time. You will not be able to do it because whenever you look somewhere, the same object will be placed on the macula of both the eyes. Now isn't that redundant? No, that's exactly what is responsible for the perception of depth. But, how does one eye know where the other is seeing? What if we have many similar objects placed around us, will our brain be fooled? This is exactly what the computer vision scientists are trying to crack from several decades. The concept is called stereo correspondence. In order to mock what our eyes are doing we use two cameras, place them at an offset similar to how our eyes are placed and take an image from both of them. When you look at such a photograph (I have a sample below) a lot of objects would have appeared in both the images, which are redundant for 2D perception but required for 3D viewing.

So these are the objects we are interested in, and need to match them in both the images to get the relative depth.



http://www.nightmare.com/~rushing/stereo/index.html

In case of our eye since the entire surrounding cannot be captured on the macula, we have to move them relative to each other to see different objects. Our retina is not a uniform sensor. In order to see something clearly we have to place it on the macula and hence the need for this movement. On the other side, a camera sensor is uniform in density and hence the entire surrounding can be analyzed just with a single shot from both the cameras, no movement required.

Your brain can actually perceive depth from these two 2D images, if viewed properly. You will need some practice for that. Here's how (http://www.vision3d.com/3views.html), if you are interested in it. To appreciate how our brain creates 3D out of these two 2D images and why we are so keen in copying from it, it's better you learn and then only proceed.

In my next post I will explain about triangulation which is the central idea behind the calculation of depth using two 2D images.

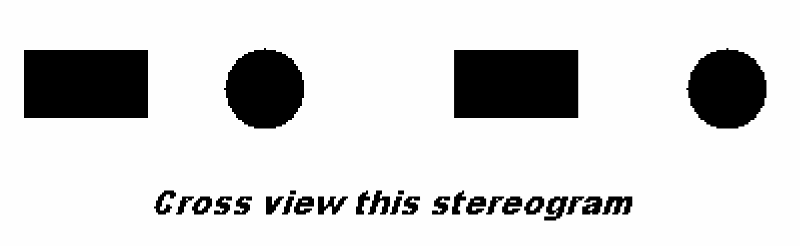
Computer Vision (11):

Practice, practice and practice till you learn to see stereograms. You might strain your eyes a lot during this process, but I bet it is worth it. If you have chosen to work in this field you have to take this up seriously. There are different kinds of stereograms available, photographs, computer generated, random dot, etc, etc. Use whichever you are comfortable with. Use smaller images initially; you will not have to cross your eyes too much. Also I feel cross stereo is much simpler to learn than parallel stereo, because crossing your eyes is easier to moving them apart (at least for me). Here are some of the links from where you can find these different kinds of stereograms.

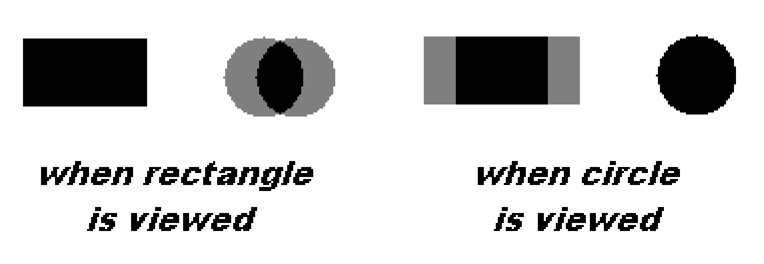
http://www.cut-the-knot.org/Curriculum/Geometry/Stereo.shtml
http://www.eyetricks.com/3dstereo.htm

http://www.focusillusion.com/YuryGallery/Yury01.php

I have created a simple stereogram here in which the circle appears to be in front of the rectangle when viewed stereoscopically (crossed).



Once you have learnt to see stereograms, there is a small observation you will have to make. You can actually do it in the above diagram itself. The gap between the circle and the rectangle is not the same in the two images. This change in the relative distance between two objects in the images is what is called disparity. Solving for depth between the two images is actually solving for this disparity. So you cannot overlap these two images one above the other to fit both the objects perfectly. When your brain combines these two images the disparity that exists between them is converted to depth. If your observation is very keen you will also be able to observe that when your brain combines the rectangles the circles would have not overlapped perfectly and when your brain combines the circles the rectangles would have overlapped at an offset. You cannot combine two objects at different depths at the same time in your brain.



The gray color in the overlapped object is shown just to highlight the partial overlap that takes place for objects at different depths other than what is viewed. Our brain does not average the colors, so you won't see this gray in the single image that your brain creates; instead it will either be the circle from the left image or the right one. This is called binocular rivalry and I want to have a separate post to explain this concept.

Computer Vision (12):

If you haven't been successful in viewing stereograms (either cross or parallel) and want to give it up, here's a simpler technique to get the same experience without straining your eyes. It is called the anaglyph technique. Here, the two photographs are overlapped before hand, so you don't need to strain your eyes to view it. Instead you use a color filter to view them. Here's a link for some examples.

http://www.rainbowsymphony.com/mars-3d-gallery.html

Go to the above link and observe the images before you move further. The concept works like this; each color filter that your glasses have should match the color component preserved in one of the two images. The single image that you see in this link is created by taking red component form one image and green component from the other and overlapping them. For example if you are using a red-blue glass combination, one of the images should have blue and not red component in it and the other should have red and not blue component. I assume that you all know a color image is a mixture of three layers; Red, Green and Blue. When you overlap the two components it will look blurry without the glasses due to the disparity present in stereo images. When you look through these glasses one of the components in the image will be filtered by each of the eye piece and so the same image will not reach both the eyes, which your brain resolves to perceive depth. It is equivalent to seeing two images either crossed or parallel. Now you know why they give you these colored glasses when you go to watch a 3D movie.

People interested in photography can take their own 3D photographs using their single 2D camera. Landscape photographers would be very excited to take such photographs, because it is very difficult to reproduce the 3D landscape effect in a 2D photo. Here are some of the ways to do it.

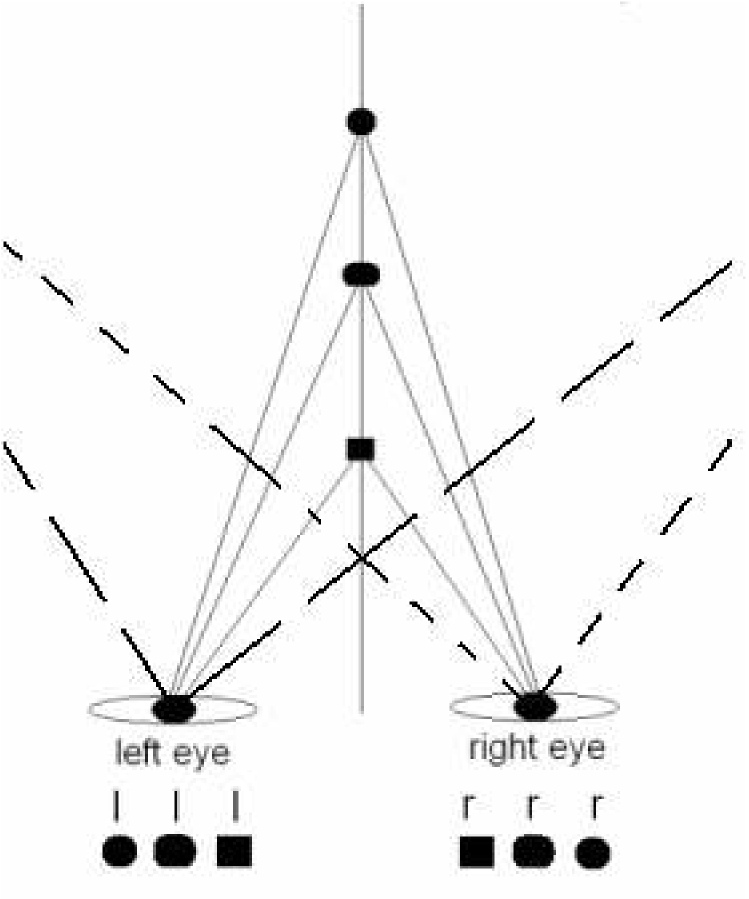
http://www.feargod.net/3dhowto.php
http://www.funsci.com/fun3_en/stscp/stscp.htm

If you find the explanation in the above links too complex let me know, I can put it in a much simpler way.

Computer Vision (13):

By now, you all must have understood that to perceive 3D from 2D image(s), the image(s) need(s) to contain similar objects with disparity. How this disparity creates the sensation of depth in our brain is by triangulation, which I will be discussing now.

The vertical line in the below diagram is the bisector between the two eyes. The square, ellipse and circle are three different objects placed at different depths from the eyes. So, what you are viewing here is the top view of the objects along with your eyes. I have not shown the movement of the eye to see the different objects shown here just to keep the diagram simple. The objects are placed on the vertical line just to get a symmetric image on the sensor and to reduce the complexity of the drawings. The dotted lines give the field of view of each eye. To get a better understanding of whatever I am trying to explain here, I suggest the reader to try these out practically as and when he/she reads through it. This will make you understand the concepts very clearly.

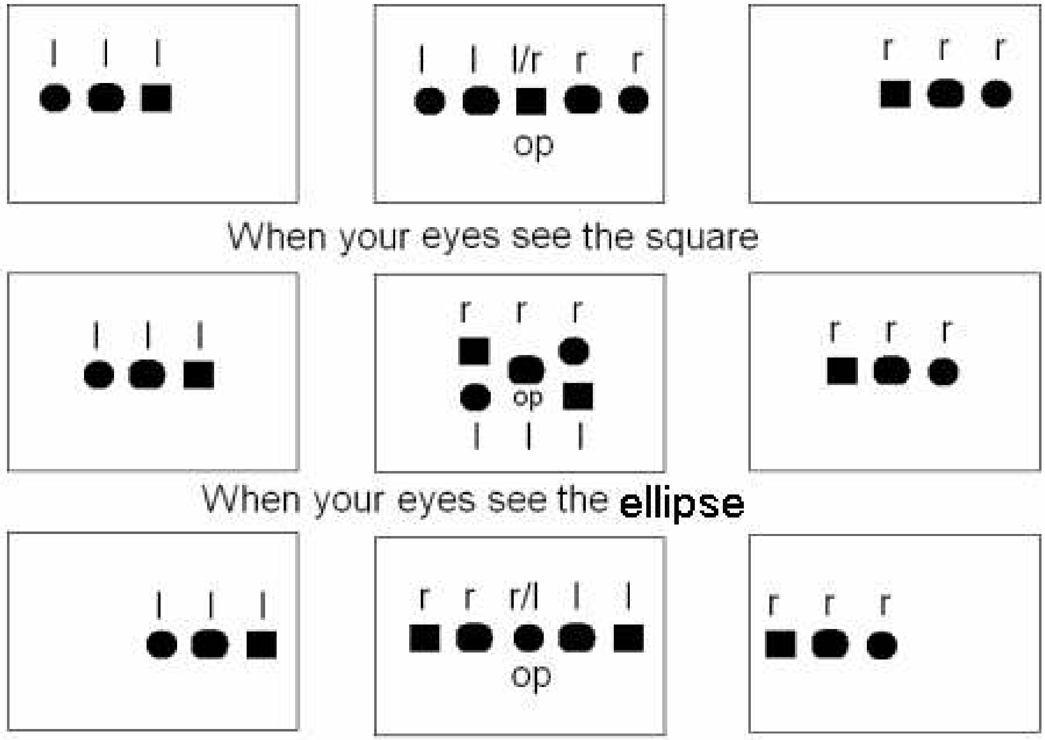


The below diagram gives the 2D projection of the 3D environment shown above, that your eyes send to your brain. For the right eye the image of the square is always to its leftmost followed by the ellipse and the circle. For the left eye the image of the square is always to its rightmost followed by the ellipse and the circle. The left column in the image below is the image captured by the left eye (objects are marked with an 'l' on top), the right column is the image formed by the right eye (objects are marked with an 'r' on top) and the central column is the combined image formed in the brain ('l' is the image that has come from the left eye and 'r' is the image that has come from the right eye). 'op' in the diagram means the overlap point, the region where the two images are combined, which in our case is the macula. Let me explain it in 3 different cases:

1. When your eyes look at the square, the square is the region of overlap in the brain and therefore the square forms the center. Other objects are moved to the sides as named in the diagram. Imagine sliding the left and the right images close to each other such that the squares are placed one over the other.

2. When the eyes look at the ellipse, the ellipse forms the center, which is obtained by sliding the two images more towards each other so that the ellipse forms the center. Here the square from the right eye and the circle from the left eye and the circle from the right eye and the square from the left eye overlap each other. They are shown one above the other for the sake of clarity. How does our brain deal with the overlap of dissimilar objects? Will it average the two or suppress one of them? This is again binocular rivalry about which I will be posting later.

3. When the eyes see the circle, the circle forms the center, which is obtained by sliding the two images further towards each other to overlap on the circle.



When your eyes see the square

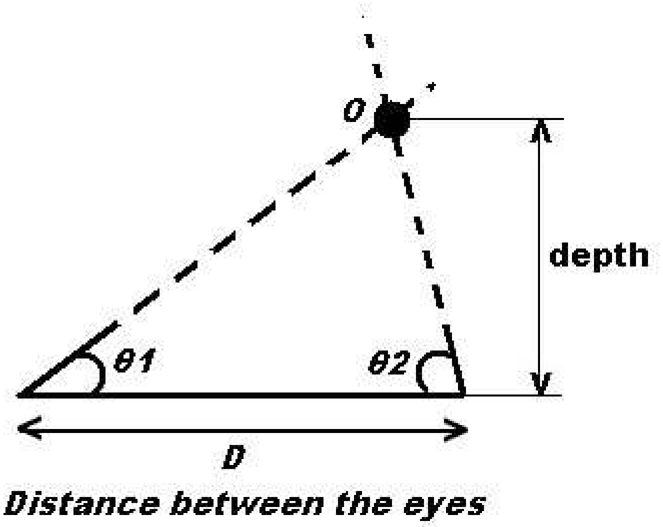
When your eyes see the ellipse

When the eyes see the circle

Try these out practically and verify the results. Ooops! This post has grown so long, which I always try to avoid but.... Anyway I will come to triangulation in my next post once again.

Computer Vision (14):

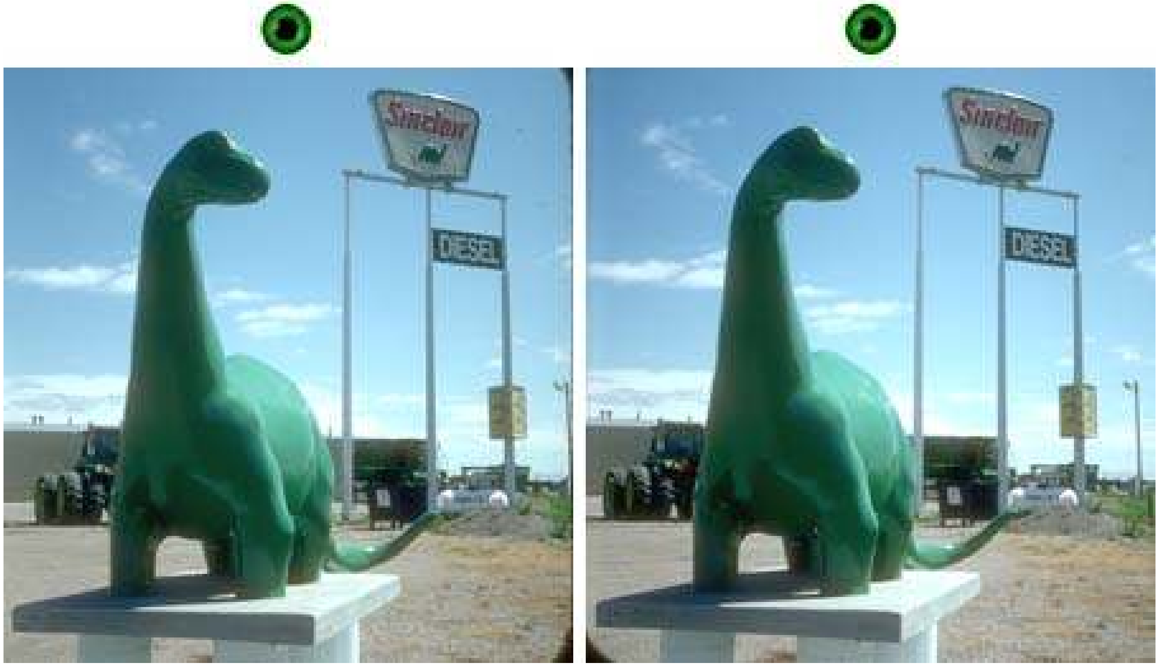
If you closely observe the first diagram in my earlier post, you will see a triangle formed whenever the two eyes see an object. The three lines that form it are; the line joining the left eye and the object it is currently seeing, the line joining the right eye and the object it is currently seeing and the line joining the two eyes. From the perspective of the eyes, they don't know where the object is, because each of the eyes has only got a 2D projection of the surrounding including the object it is currently seeing. Assuming that there is a central system controlling the movement of the eyes, this is what it knows about them. The length of the line joining the two eyes is always a constant. In order to see an object it has to be placed on the macula of the retina and hence it knows the angle at which the eyes have converged. We can easily solve for the third point which is where the object is placed and hence we know its distance, which is the depth we are trying to perceive.



Knowing the distance between the eyes (D) and the angle of view of both the eyes (theta1 and theta2), we can always extend the two lines (shown dotted) to meet at a point (O). The perpendicular drawn from O to the line joining the eyes is the depth of the object from your eyes. Since the two eyes always see a common point, the lines emanating from them always converge and make sure that a triangle is formed for an object anywhere in the common 3D space.

Computer Vision (15):

Now that we know triangulation can give us depth, how do we replicate this behavior in our computer based system? As I have mentioned earlier, the camera sensors that we use are not like our retina which has high density only in the region of the macula and less elsewhere. These sensors have a uniform density of pixels and so the object of interest need not be placed at the center of it, it only needs to be captured somewhere in its area. This eliminates the need to move the camera over an axis as done in the case of our eye. The complexity comes after the images are captured. An example of a stereo image pair is shown in the image below.



I don't remember from where I got this image, tried to search it but couldn't find.

In order to get a triangle out of an object or a point we have to find the corresponding matches of that point in the two images. This process is called **stereo correspondence**. The one reason I love this field is because it has no standards restricting you in any way. You just need to understand the problem and then you are free to come up with solutions and techniques to solve it. The problem in front of you is; for each and every point in one image how do you find the corresponding point in the other. I want to keep your minds fresh and open for new Ideaz, so I won't be detailing on the currently available techniques, because there are not one or two, but many! I strongly believe that to solve the problems of nature you just need to have an open mind to think in new ways. I want you all to give a deep thought on this problem before even trying to Google for what's already been cooked. I can assure that you still have a chance to come up with your own perfect recipe even though it's been worked out since ages.

This was the end of my introduction to "depth perception through stereo imaging". As I dive much deeper into this problem, try to think about different ways in which you can solve it. As you go on reading my posts from now on you will find that a lot of good techniques that you had thought about wouldn't really work in many cases. I will reveal the different dimensions to solving this problem along with the merits and demerits of each of them. Also open up a parallel thread and try to know what all people have been able to think of till now (You will get to know that you are not far off).

After having thought of it for so many years, I am just waiting for my brain to answer its own call!

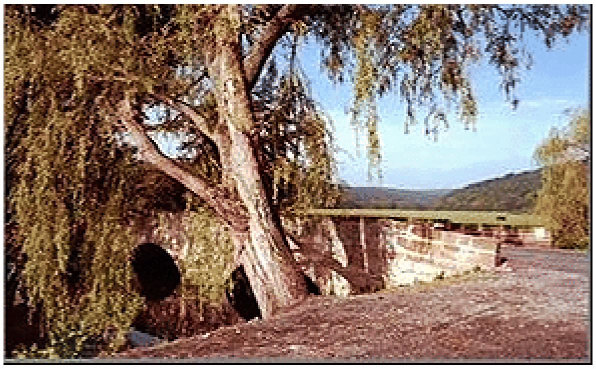
Computer Vision (16):

There is some more information left out which are not related to stereo but depth, which I want to mention before proceeding further. All the above kinds of depth perception require two images or in other words two eyes, and disparity forms the main cue to perceive depth. Such cues that the brain uses are called binocular cues. There is another category of cues that our brain uses a lot to guess depth in single images, known as monocular cues. Monocular cues are the result of the enormous amount of knowledge our brain has acquired over the years. Monocular cues help us to perceive depth in 2D images. Some links to know more about monocular cues:

http://webvision.med.utah.edu/KallDepth.html
http://ahsmail.uwaterloo.ca/kin356/cues/mcues.htm

Computer Vision (17):

Of all the monocular cues that were mentioned in my earlier post, there are just two of them that interest me a lot; motion parallax and accommodation. Motion parallax as explained in the links I had referred to, can easily be observed when you are traveling (say in a train). Objects that are closer to you appear to move faster than the objects that are farther away. If you have understood my post on triangulation, motion parallax is no new concept! Motion parallax which is a monocular cue is conceptually similar to stereovision which is a binocular cue, in the sense that both of them are perceived due to disparity. In case of motion parallax, to perceive depth along a particular direction you have to move parallel to it. When you are moving in a train, you only capture horizontal disparity between the objects, in the same way as in stereovision we perceive disparity in the direction parallel to the line along which our eyes are placed at that point of time. The first image below is a stereo image pair made into a gif and the second shows motion parallax. I don’t know how to represent gifs in a PDF, so u will have to get back to my post to view these or follow the links.



Couldn't find this image too

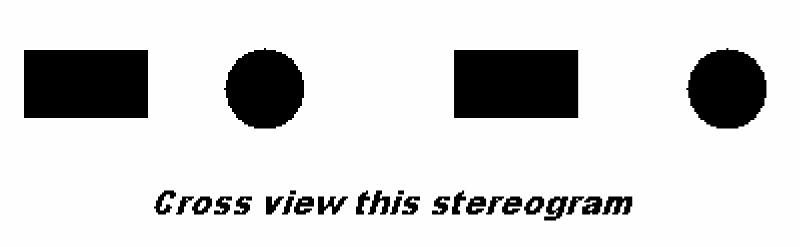
A good animation on motion parallax:
http://psych.hanover.edu/KRANTZ/MotionParallax/MotionParallax.html

Motion parallax is in fact mimicking stereovision but at two different instants of time. Imagine that instead of me, I placed a video camera and shot my train journey. If I extracted any two consecutive frames from it I would have got a stereo image pair, one taken after a small delay delta compared to the other. In the case of our eye these two images are captured at the same instant of time, while in the motion parallax case it is equivalent to moving the camera to the second eye's place to capture the second of the stereo image pair. So, when disparity can solve for depth between a stereo image pair, why not in case of motion parallax?

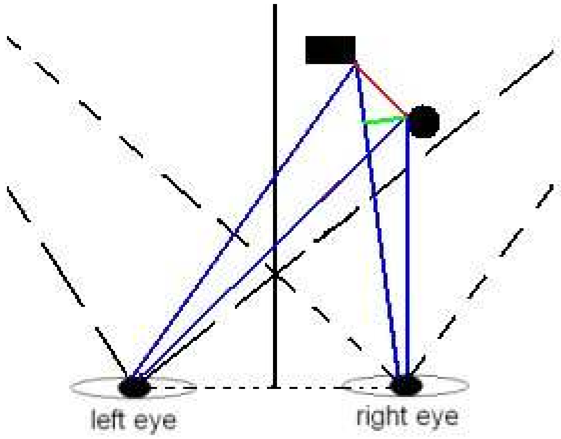
You can get more stereo images as shown above here:
http://www.well.com/~jimg/stereo/stereo_list.html

Computer Vision (18):

Disparity is something that is required to perceive depth from 2D stereo image pairs. For example, to create a stereo image pair in a computer as shown below, I just placed a rectangle and a circle one beside the other in the left image, copied the same thing for the right image as well, and then increased the distance between the rectangle and the circle in the right image.



3D interpretation of it is as follows. The image seen below is the top view of the 3D space whose 2D projection is shown above. On cross viewing it, you would see the circle in front of the rectangle. Cross viewing a stereogram means, your left eye would see the image on the right and your right eye the one on the left.



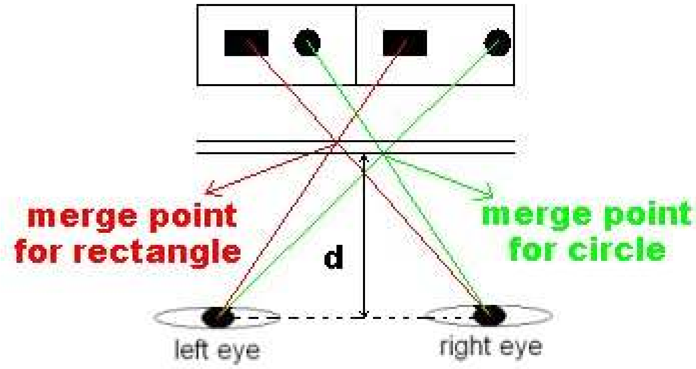
The dotted lines are the angle of view of the eyes (not to scale). The blue lines are the projection lines of the objects on the respective eyes. Since the eyes are placed at some distance from one another the projection of the objects in 3D space will always be different on both the eyes, except when the objects are on the vertical bisector. This difference in the projection lengths is what disparity is.

Disparity = length of the red line - length of the green line.

From the diagram it is clearly evident why the distance between the rectangle and the circle in the right image (red line) is kept greater than the left image (green line) to recreate this 3D effect in the brain when viewed stereoscopically. Think about how the gap should be to view the circle behind the rectangle.

Computer Vision (19):

Disparity is a must to perceive depth in a stereo image pair and so our brain needs at least two separated points with disparity to extract the distance between them. Disparity at a later stage would use triangulation to perceive depth, but this triangle would depend on the separation between the images and not the depth of the actual object. The below image illustrates triangulation from disparity when a stereogram is cross viewed.

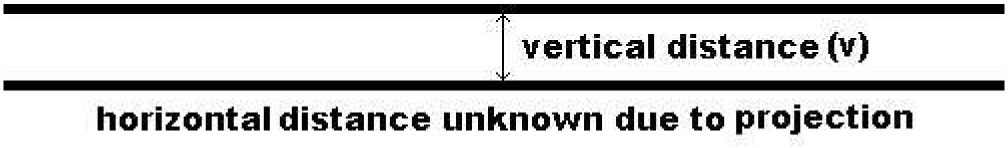


The red lines are traced when the eyes combine the rectangle and the green lines when they combine the circle. The point of intersection of the red lines gives the 3D location of the rectangle and the green lines that of the circle. As mentioned earlier the circle is in front of the rectangle when cross viewed. One of the points for the formation of the triangle comes from the point of intersection of either the red or the green lines and the other two points are the two eyes. The distance of the point of intersection of lines from the two eyes (d), depends on the separation between the images, so the absolute distance of the objects remains unknown in the stereo image pair. The relative depth of different objects from one another is obtained by corresponding objects form the two images, which moves the point of intersection of the lines according to the 3D placement of the objects (similar to red and green lines).

This is not the case when we extract depth from the actual 3D surrounding because our eye makes use of triangulation from the convergence of the eyes and not disparity. Our eyes assist to perceive the absolute depth of our surrounding while in stereograms we can only perceive the relative depth of one object from the other.

Computer Vision (20):

Even though a lot of people believe that a stereogram is exactly equivalent to seeing with both the eyes there is one major difference. Stereograms are generally shot by moving the camera horizontally by a short distance (in case of a single camera system) or by keeping two cameras side by side, which capture the horizontal disparity. Suppose there are two infinitely long horizontal bars, one at a certain distance from the other (both horizontally and vertically) and nothing else around it and you take a stereo image of this, with the camera taking the projection of their lengths, you will fail to capture the horizontal disparity, because there is none in this direction.



A camera takes the horizontal projection of objects (horizontal line pointing towards you), and so the distance between the horizontal bars along this direction cannot be shown in this 2D image. The vertical distance between them is ‘v’. In other words, if we try to capture this 3D setup in a stereo image pair to get the horizontal depth between the bars you will end up with exactly the same image in the left and right. There is no use seeing it stereoscopically, because, which point in the two images will the brain correspond? Since the camera is moved horizontally, the vertical distance between the two bars remains the same in the stereo image.

In a real scenario, how do our eyes and brain together manage to catch the right point? I mean, form a triangle and get the depth out of it. This is possible because, in addition to just 2D projection our eyes collect in real time one more parameter; focus. Focus is exactly the same as accommodation that I was describing in monocular cues. Our eye has to accommodate itself to focus (see sharply) objects at different depths. When object at one depth is seen sharply depending on the aperture of our eyes objects at other distance will be blur. This means that focus or accommodation is dependent on depth and unique for every distance from the eye. So the accommodation value would actually give the absolute depth of them object.

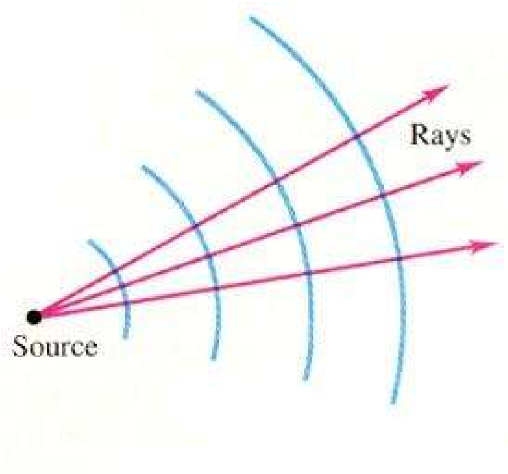
Computer Vision (21) and Optics

Let's first understand what focus or accommodation is. We have been learning since our school days that light is a ray, wave, particle, it has a dual nature, etc, etc. And lately we have also known the famous and un-debatable theory on light by Richard P Feynman et al that the dual nature of light can be explained by considering light as a particle having an instantaneous phase (which gives it the properties of a wave) associated with it. This theory is called the QED; Quantum Electro Dynamics and it promises to combine the particle and wave theories of light into one single entity that can explain almost all phenomenon of light to its highest possible accuracy. Interested people can read his own book called "QED: The Strange Theory of Light and Matter", which has his famous lectures.

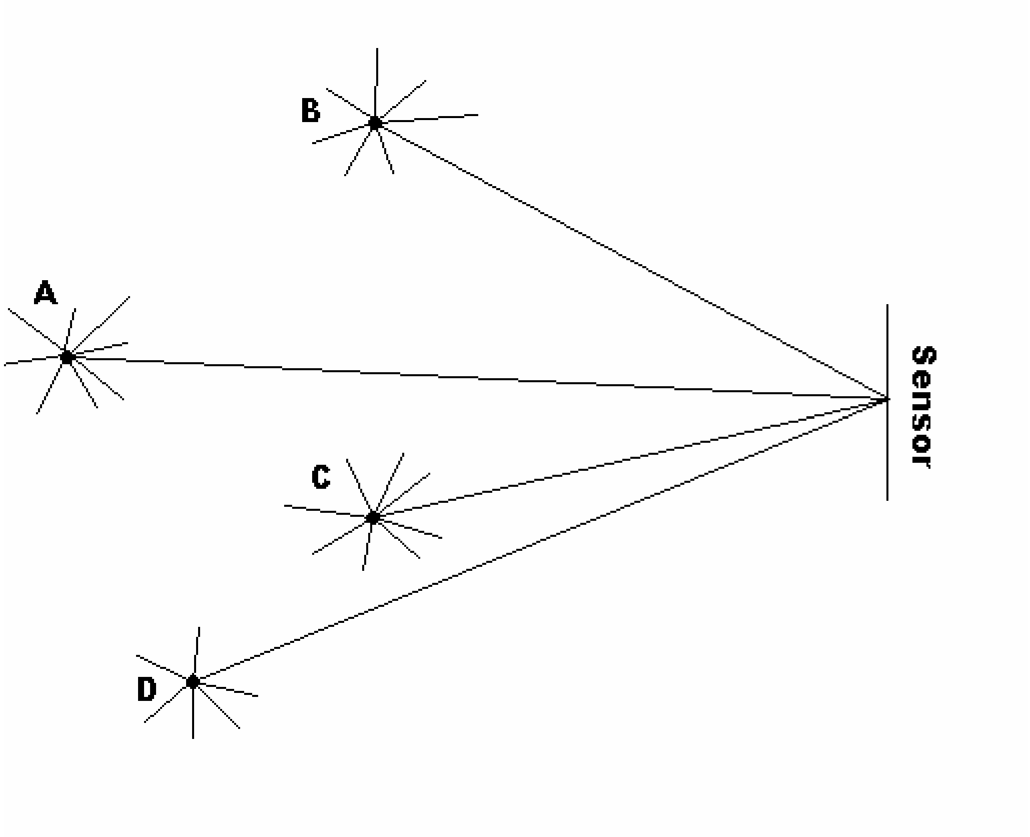
But why am saying all this? When I focus light, I am bringing together light from a region to a point and to analyze this I need to decide on the theory. For now I will not get into the complex QED, but will try to justify my experiments with simple ray diagrams. I will be coming to QED after some more posts.

Imagine that we did not have eye ball or in other words the lens of our eyes, but just had the retina to capture the light from the surrounding. How would the surrounding appear to you? There is one way to experiment this (of course not by removing the lens of your eyes :)). If you have a webcam just try removing the lens of it and switch it on exposing the retina, sorry the sensor to the surrounding. What do you see?

Computer Vision (22) and Optics



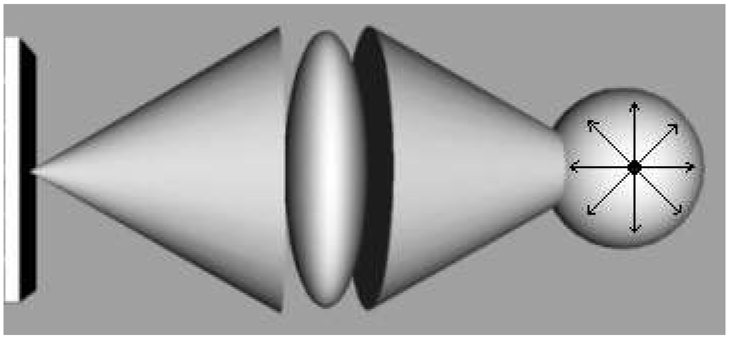
If I take a point source and place it in space it would emit light spherically in all directions around it. You will be able to see a point, only if the rays from that point reach your eyes. This means that you will be able to see a point source from any place around it. If u just had a sensor (retina) and not the lens in your eye, these rays that are diverging and almost everywhere in space would fall all over the retina to form an image which would be a uniform light patch in your brain. The same applies to non light sources as well. You will be able to see an object only if the object is reflecting light in the direction you are seeing. Again an object can reflect light in almost any direction around it. Without the lens, the reflected light from many points around you can fall at the same place on the retina as shown below.



The intensity and frequency of the reflected light from these various points can be different and hence get summed up at a point on the retina. This scenario can happen for every pixel on the sensor and hence the image that you will get will just be the summation of the intensities and frequencies of the rays coming out from various points around you. As a result of this you will always end up with a uniform patch of light on the sensor if you try to take an image without a lens.

If you didn’t have a lens in your eyes, you would only be able to know the amount of light present in the surrounding and not the objects present in front of you. The various objects wouldn’t be distinguishable at all.

To see a point as a point, we need to converge the rays that are diverging from it, to a point again. The lens does exactly this. Your brain sees various objects around it as they are because your eye lens converge the rays coming from it on the retina.

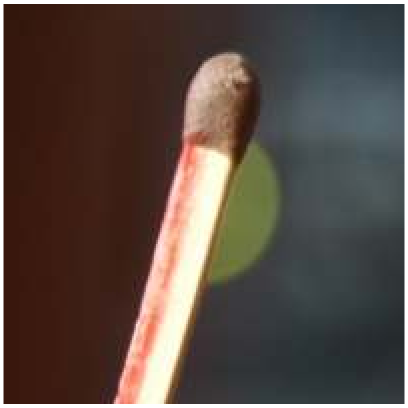


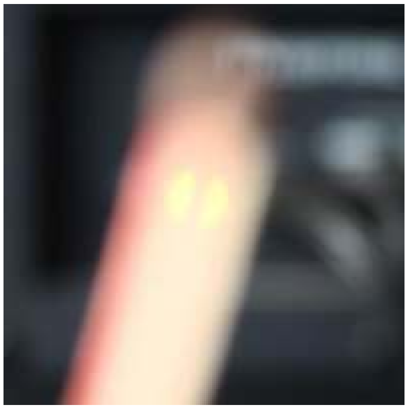
Computer Vision (23) and Optics

The best place to observe these things is in a mirror. You will be able to see any point around you at a specific place on the mirror by positioning yourself properly. This means that there are at least some rays from every point in space reaching the selected point on the mirror from where you are able to see that point in space.

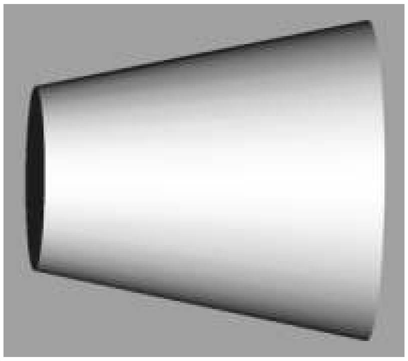
I performed a series of experiments to understand focus and the behavior of light, which I will unravel here:

SECTION1: The green light source was placed at a certain distance from the match stick. Even though the match stick had completely blocked the 2D space or projection of the light source which was a led, it is completely recovered when the focus point is shifted from the match stick to the led.





From the perspective of our eye or the camera, the light source forms a 3D cone; the apex of which is at the source itself and the base at the lens or our eye. This is the reason you see a larger circle patch of green light when the match stick is focused, which is at a distance from the led. It is like truncating the 3D cone at a particular distance from its apex. Depending on at what distance from the apex you are truncating you will be getting circles of different diameters. Larger the diameter lesser will be the intensity of the light, because the energy has now spread out.



If you take the focus point to the surface of the lens, you will see that the diameter of the circle will be the same as the diameter of the aperture of the lens.

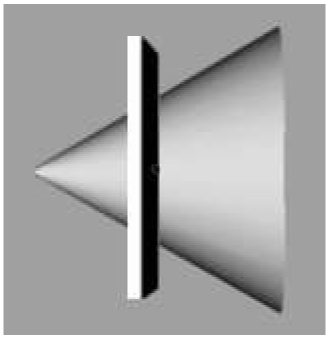
Computer Vision (24) and Optics

Even though light is traveling in 3D space, a sensor represents it on a 2D surface. Effectively what it captures is the state of light at a particular 2D plane, which is dependent on where the lens is focused. This is something that is unique; if you change the focus of your lens and the plane that you will be selecting to capture on your sensor will change automatically. Changing the plane means, selecting a plane at a different distance from the lens. This is why focus or accommodation is said to give the depth of the object when it is focused on to it.

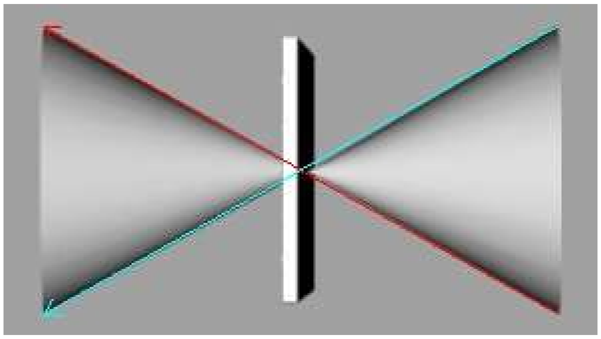
If you closely observe the three sequences of pictures I had in my earlier post you will understand it easily. In the first image the focus point was at the match stick, and the LED was at a distance behind it. The light rays diverging from this source from the perspective of the aperture of the lens would be a 3D cone which will be truncated at the matchstick. This is what is giving you that circular patch. As I move the focus back, this circle gets smaller and the intensity increases. The light that is reflected and diverging from the matchstick is now captured at a different plane, which makes it blur. Finally, when the focus point is moved to the plane of the led, it is recovered completely, even though it was masked by the matchstick completely from the projection perspective of the camera. Due to further increase in the distance of the focus point, the matchstick becomes even more blur.

Computer Vision (25) and Optics

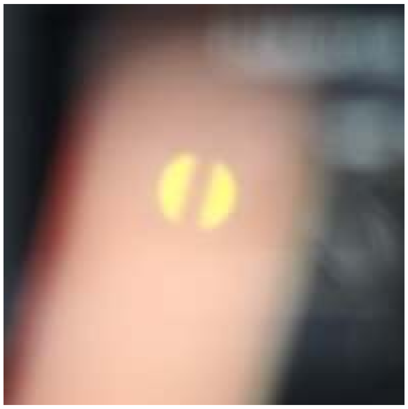
The light cone that I was describing till now will be observed when the actual focus point of the object lies beyond the sensor, i.e. the light rays from the object have still not converged when the plane of the sensor was encountered.

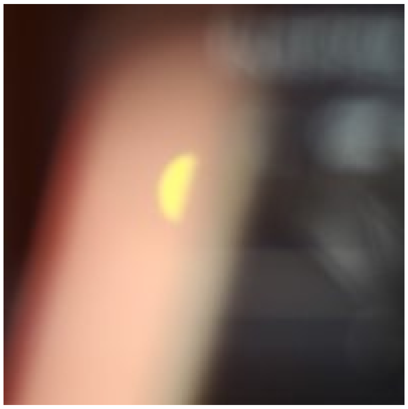


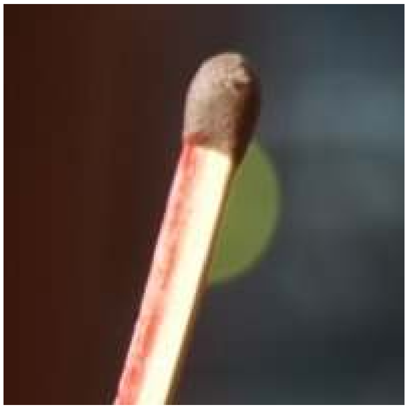
After the focus point is reached the rays crisscross and start diverging once again. Again this crisscrossing can be captured on the sensor by moving the focus point beyond the object.



The sequence of images below were taken by moving the focus point behind the object of interest; here the LED.

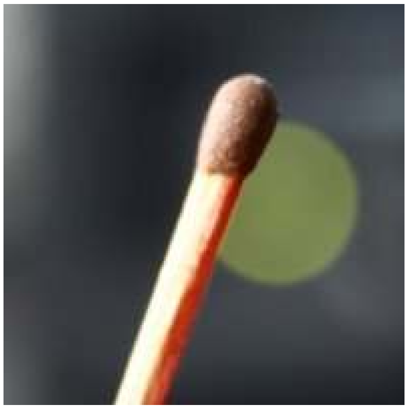


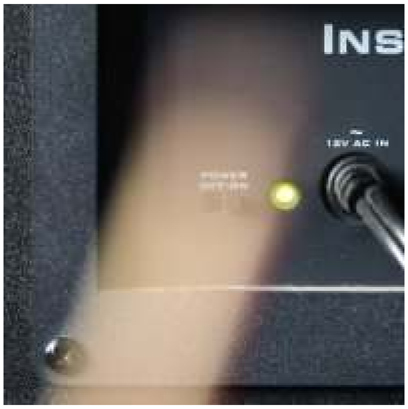


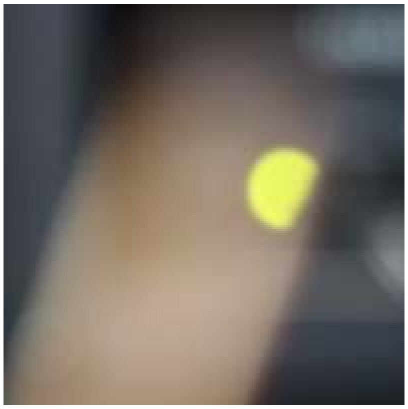


In the first image of the sequence, the focus point was moved just behind the LED and we see a similar image as when the focus point was placed between the matchstick and the LED. But now the rays have actually crisscrossed which is not observed here since the cone is symmetric. To demonstrate the crisscross nature, I placed an opaque object and covered the left half of the lens, which made the right semicircle of the circular projection of the cone, disappear! To come back to our proper cone I moved the focus point back to the matchstick and did the same experiment. Now covering the left portion of the lens masks the left semicircle of the LED! This means there no crisscross!

Computer Vision (26) and Optics



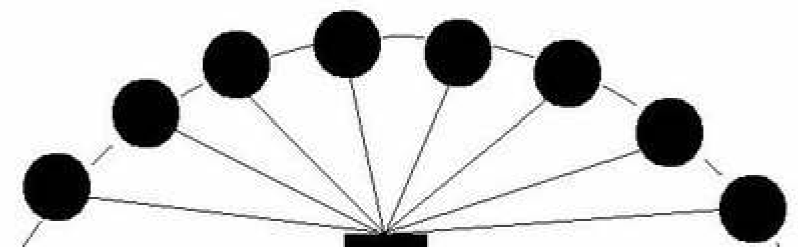




Here's another set of images that demonstrate this crisscross nature of light cones. Here I placed the matchstick at the corner and blocked any chances of light crossing the stick and reaching the aperture of the lens. You can easily find the difference between the first and the third images. The missing sector of the circle has moved to the other side, from left top to bottom right.

Computer Vision (27), Optics and Photography

The concept of the cone changes slightly, when light reflected from surfaces is taken into account. It is this light that we generally see/perceive in our surrounding, because objects reflect light and not produce light. This reflection is not the same in all the directions and hence the circular cross section of the cone will not be of uniform intensity and frequency (color). When we perceive it as a point it is the sum of all these different light rays that we are seeing.



If this sounds too complicated, just place a CD near you and try to observe a particular point where colors can be seen. From different view points you will be able to see different colors. This means that the same point on the CD is diffracting different colors. So if the aperture is big enough to accommodate all these colors, the color of the actual point will be the addition of all these. Out of focusing this point would reveal all the individual colors. One more example is the mirror, which I have already touched upon in my earlier posts. In the diagram shown above, the rectangle is the mirror and the circles are either you or your camera. Suppose you fix up a particular point on the mirror and move around it as shown in the figure you will be able to see different objects at the same selected point on the mirror. The mirror is reflecting light from different objects from the same point on it which you will be able to capture by moving around.

For all you photographers out there, bigger aperture might solve ISO problems, but depending on the aperture value you might end up getting a different color for the same pixel on your photograph. The color that your eyes see might not match the one that you get from a camera, even if you match the sensors exactly. This is because aperture also plays a role in color reproduction! Ideally you don't need a lens if the aperture of your camera is a single point, letting just a single ray of light from every point in space around it to reach the sensor. Why? You need a lens, to see a point in space as a point in the image. Normally why that is not possible without a lens is because the reflected light from objects is diverging. The lens actually does the job of converging these rays to a point, which is what focus is. When your aperture makes sure that only one ray is allowed from every point in space, there is no need to focus it! A proper image of your surrounding can be formed on the sensor without the lens. But for this to happen, your sensor should in fact be very powerful to register these single rays as a visible and differentiable value.

Computer Vision (28): Motion Detection

I know the topic on focus was too much to digest and keep up your concentration, so I decided to switch the topic a bit towards motion detection. I have only covered less than half of the full focus story, so will come back to it at a later time when I will explain one of my techniques to solve stereo correspondence.

Even though we have been able to build GHz processors and parallel computing systems, we are having a tough time matching the processing power of our brain. One reason for this is that our brain selectively processes the required information which we fail to do. On taking an image we do not know what region of the image has to be processed and so end up figuring out what each and every pixel present in huge mega pixel image can mean or form. But that's not what our brain does. It selectively puts its power in only those regions where it is required the most. For example the recognition is performed as explained earlier only in the region of the fovea. Our brain will put its concentration on the rest of the regions only when some event is detected. This event is motion. A lot of smaller creature's are specialized mainly in this kind of processing which gives their even smaller brains the power of vision.

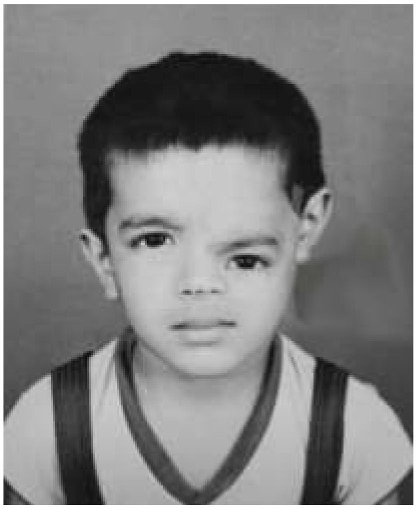
Motion detection is a concept that has been exploited in computer vision also. Then why are we still behind? When I say motion there are basically two things; motion caused due to our visual system being in motion (which involves our body motion also) and motion in the surrounding. How do we differentiate between the two? Motion in the surrounding is always limited to a certain region in space, and this small region motion detection raises an interrupt and draws the attention of our brain towards it. Motion detection is not the only thing that interrupts the brain, in fact the system that generates this interrupt doesn't even know what motion is! What it is only concerned with is whether there was a change or not. So even a sudden flash of light can interrupt your brain, even though it is not moving.

In order to detect this kind of a change in computer vision systems we try to diff the current frame with the previous one. Any change would reflect as an edge in the resultant image which would give us the location of the change. This detected change is not necessarily an interrupt to our system because we still spend time looking for this edge in the entire image. Our computer vision systems no doubt capture the surrounding in parallel at once but the processing still takes place serially, which forces it to take a back seat compared to our brain.

Computer Vision (29): Motion Segmentation

Motion segmentation is another concept that comes out of motion detection. As a newly born kid all you see around you is colors; colors that make no sense to you and you don't even know what color they are. One way in which our brain can start segmenting objects is through stereo correspondence. But again the process of stereo correspondence can be mechanical or knowledge based. If it is mechanical then we got to find how it can be done (I will discuss this later), if it is KB, our brain has to first of all learn how to correspond. So how does our brain start to segment objects? If you try to observe the point of vision of newly born kids they seem to be looking somewhere at a far off place which is the relaxed state of our eye. We need to interrupt its brain so that its visual system starts to concentrate on different things. This is the reason we get colorful toys that make interesting sounds and play it in front of them. Bright colors capture the sight of these kids and draw their attention towards it. But if these objects are placed static the interrupts stop, so also the concentration. In order to keep up the interrupts and concentration you need to keep swaying it in front of them. This not only draws its attention but also helps it to catch up on the object through motion segmentation. You can now see that its eyes are actually pointing on the object you are playing with. After repeating this procedure for quite a few times you will see that it will freeze its sight to the object even if placed static. It has now started to update its knowledge! This knowledge helps it to segment objects from its background as the days pass by and finally they will start to grasp them. This is the onset of the perception of depth.

Computer Vision (30): Why wasn't our face designed like this?



I have already touched upon the reasons behind having two sensors and their advantages (1). We now know why we were born with two ears, two nostrils and two eyes but only one mouth. What I failed to discuss at that point of time is their placement. I will concentrate more on the hearing and vision (placement of ears and eyes) which are better developed in electronics than the smell.

Our eyes as any layman will be aware of, is a 2D sensor similar to the sensors present in a camera (not design wise of course!). They capture the 2D projection of the 3D environment that we live in. In order to perceive depth (the 3rd dimension), our designers came up with the concept of two eyes (to determine depth through triangulation). For triangulation (2) to work the eyes only needed to be at an offset; horizontal, vertical, crossed anything would do. So probably the so called creator of humans (GOD for some and Nature for others) decided that they would place them horizontally to give a more symmetric look wrt our body. But why wasn't it placed at the side of our head in place of our ears? (Hmmm, then where would our ears sit? ☺)

Light, we know from our high school physics does not bend along the corners (I mean, a bulb in one room cannot light the other beside it), and from my earlier posts we know that to perceive depth our eyes need to capture some common region (the common region is where we perceive depth), they have to be placed on the same plane and parallel to it. This plane happened to be the plane of our frontal face and so our eyes came at the front where they are today.

Let's move on to hearing now! When we hear some sound, our brain will be able to determine the direction of it (listen to some stereo sound), but will not be able to pin point the exact location (in terms of depth or distance from us) of the source. That is because our ears unlike our eyes are a single dimensional sensor able to detect only the intensity of sound (of course they can separate out the frequencies, but that is no way related to direction of sound) at any point in time. In order to derive the direction from which it came our creators/designers probably thought of reusing the same concept that they had come up for our sight and so gave us two ears (to derive the direction of sound through the difference in timing when it arrives at each of the ears). To derive the direction, our ears only needed to be at an offset; horizontal, vertical or crossed, so to give a symmetric look they probably thought of placing it at a horizontal offset. But why was it placed at the side of our face and not at the top like a rabbit, dog or any other creature?

Again from high school physics we know that sound can bend along the corners and pass through tiny gaps and spread out. So you can enjoy your music irrespective of where you are and where you are playing it in your house (well! if you could only adjust with its differing intensity). So our ears never demanded to be on the same plane and parallel to it! The common signals required to perceive the direction would anyway reach it irrespective of its placement since sound can bend. Secondly our ears were required to isolate the direction in the 360deg space unlike our eyes that only projects the frontal 180deg. Probably the best place to keep it was at the side of our face.

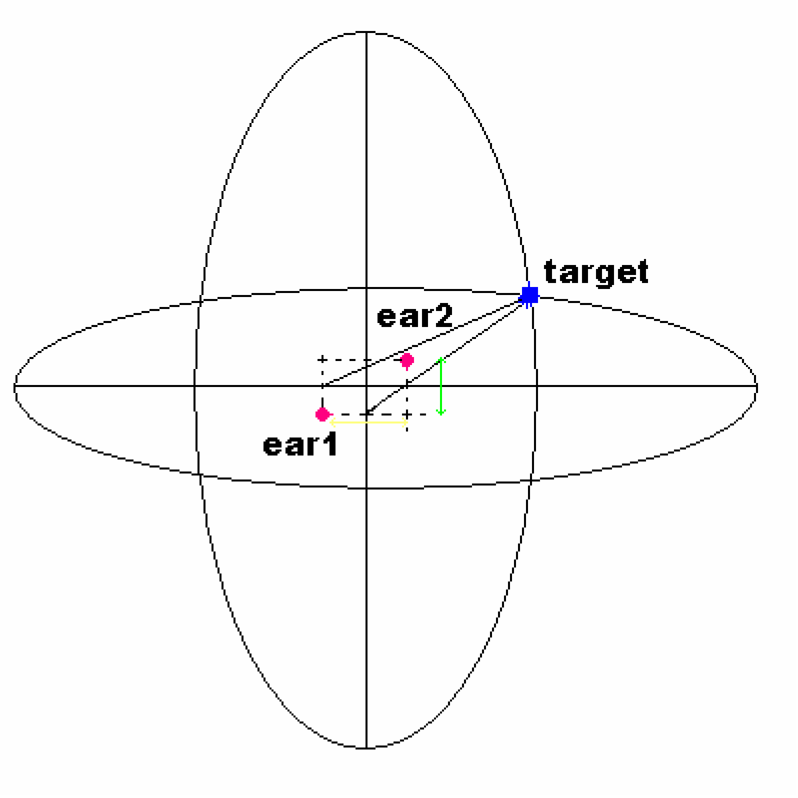
Our 2D vision was now capable of perceiving depth and our 1D hearing could locate the direction. Since our visual processing is one of the most complex design elements in our body and very well developed to perceive depth the designers never thought of giving any extra feature for any of our other sensors. But Nature has in fact produced creatures that have a different design to what we have, with ears at a vertical offset, on top of their head, etc, which I will be discussing in my next post.

References:

1. http://puneethbc.blogspot.com/2007/03/computer-vision-4.html
2. http://puneethbc.blogspot.com/2007/04/computer-vision-13.html

Computer Vision (31): “Seeing” through ears

Till a few days back even I wasn't aware of the existence of such creatures in Nature. I had not even thought of trying out something like this, even though it has been years getting into researching in this field. Nature again outwitted us in its design and complexity. I am actually talking about creatures having ears at a vertical offset to extract yet another dimension; depth that our ears/brain fail to solve through hearing. The Great Horned Owl (Bubo Virginianus), the Barn Owl (Tyto Alba) and the Barred Owl are some of such Nature's selected gifted creatures. This offset helps them to hone on a creature with more sensitivity and helps them hunt down creatures even in complete darkness. With this ability they don't even spare creatures like mice that usually hide under snow and manage to escape from their sight. Evolution has created wonders in Nature. These predators usually live in regions with long and dark winters and hence have developed the ability to "see" through their ears.



But how does it all work? With just horizontal offset our ears manage to tell us the direction of sound in the 3D space. Imagine it to be an arrow being hit in that particular direction. You don't know the distance of the target but just fire it in that direction. The arrow actually leaves from a point which is the horizontal bisector of your ears. Applying the same concept on vertical offset there will be another arrow leaving from a point which is the vertical bisector of the ears (in the case of these specially gifted creatures). From primary school mathematics we all know that two straight non parallel lines can only meet at one point in space, which in this case happens to be the target.

Even Nature can only produce best designs and not perfect ones and the Owls will definitely have to starve if their prey manages to remain silent. To make its design more reliable and worthy, Nature has never allowed a prey to have this very thought in its mind.