Digital Representation of Images

High School of Digital Culture
ITMO University
dc@itmo.ru

Contents

1 Basics of human visual perception

Computer vision systems are in many ways modeled after the human visual system. If we want to teach a machine to see the world and analyze it the way humans do, we first need to figure out how the human visual system works. How is an image formed in the eye? What kind of information can a human extract from a visual signal? What physical restrictions does the human visual system have? All of this is not only interesting but also has practical importance.

The human visual system has astonishing abilities. We recognize people and objects under very different conditions: from different angles, with different lighting, even if we see only part of the object. We can easily extract information about object sizes, their relative positions, distances between objects. Our brain can "complete" a picture if it can only see part of it.

At the same time, our visual system is far from perfect. We are prone to illusions, fail to notice many details, our perception is ambiguous and subjective, depends on our "visual experience".

Let's consider a few examples that illustrate the abilities of the human visual system.

How we see: intra-class variability of objects Take a look at the pictures in this slide. We have no difficulty understanding that all of these are pictures of different kinds of lamps, even if we see some of these lamps for the first time in our lives. Lamps can have different shapes, different colors, different sizes. But we still understand that all of those are lamps.

How we see: ambiguity of perception What do you see in these pictures? In the picture on the left, people usually see either a young woman or an old woman in a headscarf; in the one in the center, an Eskimo or a Native American; in the one on the right, a saxophone player or a woman's face. Our perception is ambiguous.

A few facts about our visual perception, 1 Our brain and our visual experience plays a big role in the way we interpret the visual signal. The brain can "complete" a picture even if we only see part of it, add semantics and recognize an object even if it's not there. People often give descriptions of an observed picture that does not match reality. All of us can recognize "something" or "someone" in the shape of a cloud.

The way our brain completes and interprets a picture depends heavily on our visual experience. Our visual system learns throughout our entire lives. We are better at recognizing familiar images. For example, a typical European finds it difficult to recognize faces of Asians, while people who grew up among Africans find that all white people "look the same".

Look at this black-and-white picture. Can you recognize what it is? What if we turn it upside down? In its familiar orientation, it is much easier to recognize the shape of a world map.

A few facts about our visual perception, 2 A couple of illustrations confirming that we often look for familiar entities in a picture and that our brains “complete” pictures. These are examples of well-known optical illusions where the eye adds non-existent information.

In the picture on the left, an outline of a square is clearly visible, despite the fact that the image is missing the lines defining that shape. A similar effect involving a circle is seen in the picture on the left. Just a few lines are enough to produce the illusion of a circle.

A few facts about our visual perception, 3 Where we grew up determines how we “see” the world around us. For example, do the steps in this picture go up or down? Most of those of us who reads and writes left to right would say that they go up. For those who are used to reading right to left, the steps go down.

The structure of the human eye The eye has an almost spherical shape with an average diameter of approximately 20 mm. The eye is basically an optical sphere. Light comes into the eye through the pupil, which is a hole in the center of the iris, is refracted by the lens, and is projected onto the eye’s retina.

The pupil can constrict or dilate, thereby regulating the amount of light passing in. The pupil’s diameter can change between 2 to 8 mm.

The lens is located behind the iris and the pupil. Unlike a hard optical lens, the optical power of the eye’s lens can change due to its changing shape. As the picture shows, the curvature radius of the front surface of the lens capsule is larger than that of the back one. The lens’s shape is changed by loosening or tightening the front and back portions of the ciliary zonule fibers. To focus vision on a distant object, the ciliary muscle relaxes, the choroid contracts, pulling on the ciliary zonule fibers, flattening the lens. To focus on a close object, the ciliary muscle contracts, which makes the lens rounder and increases its refractive ability. As the lens’s refractive ability changes from its minimum to its maximum, its focal distance changes from 17 mm to 14 mm. When examining objects further than 3 meters away, the lens’s refractive ability is minimal, and when examining close objects, the lens’s refractive ability is maximal.

When the eye is properly focused, the light from an external object is projected in the form of an image onto the retina. Across the inner surface of the retina, there are discrete light-sensitive cells (receptors), which record the image. There are two kinds of receptors: cones and rods. Cones are highly sensitive to light’s spectral components and ensure photopic vision, or vision in bright light. There are 6 to 7 million cones in a human eye. They are generally located in the central part of the retina, called the macula lutea. In the center of the macula there is the fovea centralis: the area of most accurate vision. Humans distinguish fine details of an image primarily due to cones. Each cone is connected to a separate nerve ending. The eye’s external muscles rotate the eyeball so that the desired object's image falls within the macula.

There are a lot more rods in the eye than cones: between 75 and 50 million. Unlike cones, rods are spread all over the retina, and each nerve ending connects to multiple rods at once (about 10 on average). This decreases their ability to distinguish image details. Rods can form a general picture of the whole field of view. They are most sensitive with low lighting levels and do not participate in color vision. This is why objects that are brightly colored under bright lighting look like colorless shapes in dim lighting, since only rods are activated in dim light. This is known as scotopic, or twilight, vision.

Brightness adaptation and contrast sensitivity, 1 Knowing about the eye's ability to distinguish different brightness levels has great practical importance. It lets us choose the correct method for discretizing brightness levels for representing digital images.

The human visual system can adapt to a huge range of light intensity levels, on the order of 1010: from the threshold of scotopic vision to the glare limit. Experiments also indicate that subjective brightness (i. e. brightness as perceived by humans) is a logarithmic function of the light intensity that reaches the eye. The slide shows a graph of light intensity versus subjective brightness. The long solid curve represents a range of intensities to which the human visual system can adapt. In photopic vision alone, this range is about 106. At low intensity levels, there is a gradual transition from photopic vision to scotopic vision.

Brightness adaptation and contrast sensitivity, 2 But it's important to understand that this is dynamic range: the human visual system cannot operate in this entire range simultaneously. It accomplishes such a large variation by changing its overall sensitivity. This phenomenon is known as brightness adaptation. The total range of simultaneously distinguished intensity levels is relatively small compared with the total adaptation range. For a given set of external conditions, the current level of visual system sensitivity, called the brightness adaptation level, corresponds to a certain brightness, for example, the Ba point on the plot. The short intersecting curve represents the range of subjective brightness that the eye can perceive when adapted to the specified level. This range is rather limited: all brightness levels below Bb are subjectively perceived as black and are indistinguishable. The upper dashed part of the curve is not limited, but loses meaning at some point, because an increase in brightness simply results in an increase in the Ba adaptation level.

The ability of the eye to discriminate changes in light intensity at any specific adaptation level is also of significant interest. A classic experiment used to determine the capability of the human visual system for brightness discrimination is as follows. A subject is asked to look at a large flat uniformly illuminated screen. The screen should be large enough to occupy the entire field of view. The screen's brightness (marked as I on the slide) can be varied. To this uniformly illuminated