Visual Perception and the Aesthetics of Photography

[Distributed to Future Generations Anonymously]
November 18, 2004
Introduction to Cognitive Science
Professor Brian Scholl
Art and the experience of beauty seem far removed from a materialistic, computational approach to how the mind works, and even from science in general; however, cognitive science can teach visual artists important principles of aesthetic perception that can help them maximize the impact their art has on viewers. Specifically, when cognitive theories of visual perception and attention are applied to the art of photography, it becomes clear that a good photograph must be a self-contained, ordered representation of space, and that it must direct the attention of the viewer to objects that elicit strong associations in his or her mind. Cognitive science can also describe the nature of the challenge the photographer faces when trying to record a visual scene in this way using a simple tool such as a camera.

Before analyzing the making and viewing of photographs, it is necessary to examine some of the principles of visual perception in general. Essentially, the task the visual system carries out is to make an ordered representation of three dimensional space from the pattern of incident light hitting the retina. The viewer first orients to the most salient stimulus so that it is near the middle of the visual field, where the fovea is, so that detail is seen the sharpest. The visual system then looks for edges and lines to delineate and parse objects, it uses shading to estimate the shape of the objects, and higher-level processes recognize and label the object and then bring up the relevant associations to make the appropriate response.

This higher-level information is useless, however, without a sense of where these objects are relative to the eye and to each other, and how they are moving. The visual system uses several mechanisms to judge depth, the first being calculations on disparity between the images that hit each retina. The closer an object is to the eyes, the greater the difference between the images it casts on the retinas. The second mechanism of depth perception uses focus. In order to see the sharpest detail in part of a scene, muscles in the eye contract and relax so that objects in a particular plane of space (that is parallel to the plane in which both eyes are) become in focus. Objects that lie beyond this plane and object too close to the viewer will be out of focus; therefore, the relative level of focus of an object is a measure of its distance from the viewer. The third mechanism involves the shape of the image that objects cast on the retina. An object that is far away will cast an image onto the retina that is smaller relative to the image cast by an identical object that is much closer. This also affects the perception of single objects that have parts at varying distances from the eye. For instance, a rectangular object will not cast a rectangular image onto the retina if it is extending away from the viewer. Instead, the end
furthest away will appear smaller than the closer end, so the sides will no longer be parallel, but will be bent towards each other. The visual system understands these principles of optics and, together with the information gathered from the binocular disparity and the effects of focus, uses them to calculate how far away objects are from each other and from the viewer, and how they are oriented.

Often times, however, the exact position or structure of something can not be comprehended from a static image alone. This is because in some instances one retinal image could result from more than one stimulus or spatial configuration. But if the object and the viewer move relative to each other, the object can be seen at varying angles, and the resulting information on how the retinal image changes with movement can be used to reduce the ambiguity. A related function of the visual system is that if the objects are moving too quickly relative to the eye, they appear blurry. This is because as the image of an object moves across the retina, it activates a quickly changing set of retinal cells, and as it moves into a new area it stimulates new cells before the previous ones return to the stable firing rate. Therefore, objects that speed by appear to be occupying more than one position in space at one time. This is a byproduct of the neural functioning that has an interesting effect on the perception of photographs, as will be described later.

In sum, the visual system takes in a tremendous amount of information from the images cast onto the retinas and then breaks them down into distinguishable objects and calculates their shapes, sizes, positions, and motion. The end product is a sophisticated three dimensional map of the space, with a representation of the types of the surrounding objects. It is important to avoid thinking of the output of the visual system as a simple reflection of the visual stimulus presented to it, because what is actually produced is an ordered representation constructed from the chaos of the electromagnetic waves bouncing off the world, and this representation reflects just as much about the structure of the visual system as it does about the incoming light.

It turns out, though, that at any given moment, the viewer is acutely aware of very little of this information. “Given the richness of our visual world,” say Simons and Levin (1998), “it is perhaps unsurprising that we cannot represent all the visual details of every object and instead must focus on a few important objects” (644). The brain simply can not keep track of every part of the visual scene. Therefore, it attends to the objects that are relevant to the task at hand and to the thoughts being processed in consciousness at that moment. The rest blends into the
background and becomes less noticed. There is even evidence that objects that are not relevant to
the current task will be actively ignored or repressed (see Most et al. (2001))². But if something
especially important or salient enters the scene, attention will shift to it. So, one could theorize
that there is a loop of information that flows between the highest level processes of
consciousness and the visual system: the content of awareness and conscious thoughts determine
what things are visually attended to, but the visual system also inputs new information into
consciousness and changes the flow of thought.

This relationship between conscious mental states and selective visual attention, and
especially the effect the former has on the latter, will become crucial to understanding the
content of photographs, as we will see. This concept is perhaps best explained with a simplified
example: consider a person hopelessly lost in the jungle. He will be focusing on finding passable
paths through the trees, getting back to the road, and avoiding injury and encounters with
threatening organisms. Therefore, the set of objects visually focused on during this period will
probably include stones on the ground, trees in the way, the shape of the land on the horizon,
sticks and roots that look like snakes, actual snakes, and the leopards that leap out from behind
the trees. This set would probably exclude the shape of the leaves of the trees, the plumage of the
birds perched in foliage, the clouds, and most stationary objects directly not directly in front of
him. Therefore, the set of visual stimuli he is aware of depends on his state of mind and the
problem he is thinking about, and these objects are related to each other because they are all
important for safe navigation through the environment.

Now that we have established how the visual system makes sense of the world and
provides the viewer with an awareness of what is going on, we can turn to photography, which is
a specialized use of the visual system. The photographer uses his visual system to seek out
moments in time when space is arranged in a beautiful, meaningful way, and then he tries to
record these moments for other people to experience. But in order to do this, he uses a camera
instead of his own brain, and in order to do it successfully, he needs to understand the similarities
and differences between these two tools.

The camera is in many ways like an eye: it has a lens that projects incoming light onto
light sensitive material at the back; it can focus by changing the distance between this material
and the lens; and it has a round opening that can expand and contract to let in varying amounts of
light. But the similarities between the camera and the visual system as a whole end here. The
camera carries out no computations, it doesn’t recognize objects, label them, and attach meaning to them, it doesn’t extract three dimensional space from the two dimensional piece of film, and it doesn’t selectively attend to the most important objects in the scene. Rather, it takes all of the incident light in with all its chaotic lines and curves and colors and flattens it all out onto a piece of plastic coated with silver particles. After that, this chaos is frozen forever in its pure form.

This presents the photographer with significant challenges and also provides him with opportunities. The camera can do very few of the functions that his visual system carries out to create order from the disorder of the scene, so he has to artificially replicate these processes with the camera, or else somehow compensate for them. If this is done, the resulting image will reflect what the visual system tries to extract from the world - namely, order - and the viewer will be able to understand what the photographer saw in the scene, which is more than the brainless camera would see on its own.

First, the photographer must compose the picture so that the important objects are placed in the frame similarly to the way an observer would place them in her visual field. Salient objects will be towards the middle, and nothing important should be cut off by the edges of the frame. If only a bit of something is sticking into the picture, the observer will notice it and find it aggravating, because in normal perception she would turn her head to see all of it, but since the frame of the image is fixed, that can’t be done. A second factor the photographer must keep in mind, which also results from the fact that the picture is frozen in time, is that normally people judge the structure of an object by moving relative to it to observe it from various angles. Since in a photograph the perspective is fixed, the photographer must make sure that this angle is the right one, and that the objects are sufficiently comprehensible.

Third, the photographer must recognize that the camera does not have the advantage of having two eyes, so depth in a picture can not be judged from binocular disparity. The distances of objects must be inferred from the relative size of objects at different distances and from the effects of focus. This latter method is tricky because the depth of the plane of focus can be altered by changing the size of the camera’s aperture. On the one hand, a small aperture will bring into focus all objects in the foreground, middleground, and background. The scene will appear very flat, as if objects in the foreground are very close to those in the background. On the other hand, a large aperture will make the plane of focus very thin, so that if the photographer focuses the camera on an object in the middleground, the foreground and the background will be
blurry. This provides the opportunity to exaggerate depth, and in turn to skew with the perception of the size of objects, which is related to the perception of how far away they are.

Fourth, it is useful for the photographer to understand that blurry images are interpreted by the visual system as resulting from motion. Specifically, blur is only caused by the motion of images across the retina at a certain threshold speed. On film, blur can be caused by motion of just about any speed, depending on how long the shutter is open for. Therefore, the visual system will automatically assume that the blur in an image came from frantically high speeds, which can be misleading. But depending on what the photographer is aiming for, this may or may not be a drawback.

Finally, the clever photographer needs to keep in mind that the human brain can only separate out and attend to a few items at once. If there is too much chaos, if the objects overlap each other, are too numerous, or too close together, the visual system will give up trying to focus on individual aspects of the scene and just blend it together into one area of unimportant disorder. However, this does not mean that there can not be more objects in the picture than can be attended to at one time, because inside the viewer can visually scan the photograph and notice different things in sequence. Therefore, there needs to be a balance between overwhelming confusion and simplicity, and each of the important objects needs to be salient and separable from the scene. The order in which objects are attended to will depend on their size, sharpness, contrast to the surroundings (in terms of color and luminosity), and perhaps even on the viewer’s belief about how likely an object is to be found in such a scene. With careful thought, the photographer can control this, and by doing so he can affect how the viewer thinks and feels about the picture.

The number of important objects in the frame, the arrangement of these objects, and the extent of focus and blur are the most basic parameters that a photographer should keep in mind when trying to make a visually appealing picture. The above suggestions do not represent absolute laws, but in general they allow for the most basic perception of beauty. The person looking at such a photograph will feel, in some abstract way, that it is good. This is because the photograph replicates the way in which the human visual system is constantly striving to see the world – in terms of complete representations of discrete parts of space in a particular time. Such photographs will have an overall sense of order in them, and this is what our perceptual systems seek out over everything else.
In order to do all of these things, the photographer must cure his own instinct blindness. He has to carefully replicate the functions of his nervous system when using the camera, which is essentially an eye without a brain, because normally these functions are carried out effortlessly and automatically. He doesn’t even know that his brain is doing this much processing on the light that hits his eyes, let alone how it works. As a photographer learns to take good pictures, he will slowly notice that the camera doesn’t see the same things he does, so he will slowly figure out how to use the camera to construct an image that represents the representation of space that he has in his mind. This is where an understanding of cognitive science can help him. If he knows precisely what processes are going on in his brain, and how they work, he can more quickly learn to replicate them with the camera.

He probably suffers from instinct blindness in the first place because the processes that have been mentioned above are encapsulated in domain-specific mental modules. The depth perception module, for instance, selectively takes in information about the relative sizes of objects, their relative levels of focus, and the disparity between the images they cast on the two retinas. The module quickly processes this information and then sends depth values up to other processes; i.e., the center that constructs a map of three dimensional space. Other parts of the brain do not have access to the computations that go on in this module, so it is said to have “informational encapsulation” (Carston, 62). When taking a picture, the photographer must become aware of how these computations work so that he can control how the viewer will perceive depth in the picture. This same idea applies to the recognition of objects based a representation of their shape and the selection of which salient items to attend to and then include in the photograph.

The issue of the selection of subjects leads us deeper into the aesthetics of photography, where it is not just the geometrical properties of the picture that matter, but rather the strength of the effect it has on the viewer’s thoughts. William James hypothesized that conscious thought consists of a string of momentarily stable cognitions linked together by the fleeting activation of a network of associations. The stable thoughts make up our consciousness, and James called them “the nucleus.” The network of associations finds links between the present thought and the one that follows it. This network is almost completely outside of consciousness, so James labeled it “the fringe.” A cognitive scientist named Russell Epstein suggests that the faint sensation of the activity of the fringe is related to the perception of beauty. “The sense of beauty we feel
when we encounter a beautiful object,” he says, “is the sense that it has a deeper meaning that is not entirely expressed in its surface features … I postulate that the "something of a quite different kind" that underlies these sensations is the same network of associations that underlies the stream of thought” (225). But how does this relate to the creation and appreciation of photography?

As described earlier, in normal day-to-day vision a person perceives things terms of a constantly changing set of salient visual stimuli. This set of objects that are visually attended to is part of James’s nucleus. They are connected to each other, just like all components of conscious thought, by the associations that make up the fringe. When the photographer sees something to take a picture of, there is a strong sense of association, of meaning, in the visual scene; between the objects themselves and between the other beliefs and ideals the photographer has in his mind. For instance, a certain visual stimulus may be associated in the mind with ideally beautiful things, or it may be so remarkable that it is linked to other remarkable phenomena that the photographer feels need to be recorded. Either way, photographer is out hunting for beauty, which means he is waiting to see connections between things that make the things part of something larger. When he finds it, he tries to transfer this mental state onto a piece of film.

If he does this according to the principles described earlier, something remarkable will happen when a viewer looks at the picture. The photograph will be constructed in such a way that brings to his attention to the same set of objects that the photographer noticed. Since the visual system normally attends to objects that are linked together or to some deeper concept, his brain will search out these links. If a strong network of association is activated in his mind as a result of this, he will find the photograph beautiful. He will have gathered information from the picture that is not written the pattern of light and dark that makes up the photograph, but is solely in his mind and the mind of the artist. We feel beauty as a form of pleasure because we are presented with a literal representation of reality in which things fit together, and in which there is order on every level. In a world where finding order is a constant struggle, this is cause for joy.

Cognitive science reveals that there are two parts to the beauty of a photograph. First, there is the basic sense of order, unity and clarity (although not necessarily simplicity) that arises when the geometric qualities of the photograph meet the basic requirements of the visual system. The photographer accomplishes this by overcoming his instinct blindness in order to replicate the
functions of his visual system for his brainless camera. Second, the more profound sense of beauty results from the activation of a nexus of associations in the mind of the observer, and this nexus represents concepts beyond the scope of the image itself. This phenomenon results from the fact that the brain does not randomly choose which visible objects to attend to, it makes this choice based on the abstract relationships they have with each other and with other mental states. If the photographer knows how the visual system works, he can present this same set of objects in a photograph so that analogous associations are made in the mind of the observer, who is then moved by the deep sense of order in the composition of the image and the meaning of the subject matter. The photographer, therefore, whose goal is to inspire the appreciation of beauty in other people, can benefit from the lessons of cognitive science. But to the person looking at the photograph, the principles of cognitive science matter very little, because beauty is perceived automatically, with hardly any thought, and this is precisely why it is so magnificent.

Sources:

4 As quoted in Epstein (2004).