Vision Science for Display Technologies

Shin'ya Nishida[1,2]

nishida.shinya.2x@kyoto-u.ac.jp
[1] Graduate School of Informatics, Kyoto University, Japan
[2] NTT Communication Science Laboratories, Japan

Keywords: Human vision, Perception-based display, Image-computable model

ABSTRACT

By leveraging vision science, we have found innovative solutions for realistic projection mapping and 2D/3D capable stereo, and have gained new insights into the perceptual effects of display frame rate and display gamma. Image computable models of human vision will further advance perception-based display technology.

1 Introduction

The scientific understanding of the principles of human visual processing is an exciting research goal in itself. In addition to that, the properties of the visual system revealed by science have advanced display technologies. Here I will first introduce our display systems leveraging human perceptual characteristics, and then explain perceptual mechanisms that may contribute to the future development of display technologies. Finally, I will explain a promising strategy that I believe to accelerate cooperation between vision science and display engineering.

2 Projection mapping that makes a stationary object appear to move

Some may believe that the complete physical reproduction of visual information that reaches the human eyes is the ultimate goal of display technology. However, instead of seeking physical reproduction, perceptual reproduction that takes into account the processing characteristics of the human visual system can sometimes be more efficient while maintaining perceptual quality and fidelity and can realize functions that could not be achieved with physical reproduction. What follows are two example systems we have developed.

Deformation lamp (Hengentou) is a projection mapping technology that makes stationary objects appear to move [1][2]. The algorithm is quite simple. Capture an image of a target stationary object by a camera. A movie is created by adding appropriate motion to the captured still image. A differential image movie is then created by subtracting the original still image from the movie. The monochrome (grayscale) version of the differential image movie is projected onto the original still object with a precise image alignment. This is all we need to make a stationary object appear to be moving.

Just by hearing this principle, you may consider that our technique cannot work properly, since it does not produce correct motion (displacement of the pattern) in the target image particularly about the color components. However, humans perceive motion when the motion detectors in the brain are properly activated. Knowing that the motion detector is designed to detect the local intensity flow in each sub-band of the spatial frequency decomposition (wavelet transform), one can understand that the observers can perceive realistic motion only by projecting the grayscale differential movie. Furthermore, knowing that the visual system processes color and luminance separately to some extent, that motion perception mainly uses luminance information, and that when there is a discrepancy between color and luminance motion, the color component is captured by the nearby luminance motion component, one can understand that color appears to move together with luminance motion even when only luminance motion is projected. Furthermore, since the projected light is a grayscale image aligned with the original pattern, the observer is not likely to notice the projection, and the stationary object appears to be realistically moving. Because our technique does not require physically accurate pattern reproduction, it can be used even under bright ambient lighting. In this way, by leveraging human perceptional characteristics, Deformation lamp provide a simple and powerful augmented reality method with a wide range of applications.

3 2D/3D compatible display without image blurs even when 3D glasses are not used

Hidden Stereo [3] is an extension of Deformation lamp technology for binocular 3D stereopsis. The same disparity induction pattern is added or subtracted from a single image to produce left and right images. As a result, a stereoscopic 3D image is perceived when viewed with stereo glasses, while a clear 2D image is perceived when viewed without stereo glasses since the left and right images are perceptually fused and the original image is restored. This is a 2D/3D compatible display technology that allows users to enjoy both 2D and 3D by just taking off and putting off stereo glasses.

The key point of this technology is how to create a disparity induction pattern. It is known that each of the binocular disparity detectors in the human brain has a preference for a specific position/spatial frequency/orientation, and computes the binocular disparity (depth) based on the phase difference between the two eyes. To make this phase difference, we use a

90-degree phase-shifted image. When the phase-shifted image is added to the original image, the phase of the composite image is shifted by 45 degrees if both images have equal contrast. By changing the contrast of the 90-degree phase shift image, the amount of phase shift of the composite image can be controlled. Using this principle, a disparity induction image is created to achieve the desired pattern of the binocular disparity (half of the disparity required for 3D presentation) by adjusting the local contrast at each position, spatial frequency, and orientation. The same disparity induction image can be subtracted to create an image for the other eye with the direction of phase shift reversed. By presenting these images to the left and right eyes, the observer can perceive a 3D scene. Although the range of depth that can be reproduced by this technique is limited, one can experience a high-quality 3D image when the depth range is appropriately adjusted.

The stereo image pair generated by Hidden stereo is physically (geometrically) incorrect, but it drives human binocular disparity detectors as do the normal stereo image pairs. Therefore, although Hidden stereo achieves 2D/3D compatibility that cannot be easily achieved with physically correct stereo systems, the human observers find little problems when seeing Hidden stereo.

4 Trajectory integration and temporal properties of the visual system

Next, I will introduce two of the processing characteristics of the visual system revealed in our research that may be deeply relevant to display technology.

The human visual system cannot utilize 100% of the physical image information. It can detect only a limited range of the physical image signals. Its limitations and characteristics have been used to specify the basic principle of display design.

For example, there is a limit in the spatial resolution, and fine patterns beyond 30 cycles-per-degree are almost invisible (when visual acuity is 1.0 or 20/20). This spatial frequency characteristic is an important factor that determines the display resolution of a display. The spatial resolution limit can be ascribed to such factors as aberration, diffraction, and cone density.

Similarly, there is a limit in the temporal resolution. The temporal frequency response of the visual system has been psychophysically measured by the limit of flicker detection for rapid alternations of light and dark or different colors. The upper limit of the flickering sensation is ~40Hz for luminance and ~15Hz for color. In other words, visual sensory signals are integrated within a time window of ~12ms for luminance and ~30ms for color.

The visual temporal limit indicates how fast the display should be updated to suppress undesired flicker. To present temporal modulations up to 30Hz, the display frame rate should be 60 Hz to produce light/dark flicker, and 120 Hz to produce unambiguous sinusoidal motion (90-degree phase shift).

Traditionally, it has been considered that the critical fusion frequency is a characteristic of the temporal filter (low-pass or band-pass filter) in the visual system and is determined by the early-stage hardware. Since the stimulus changes faster the critical fusion frequency have no impact on the visual processing following the initial filters, if a display accurately presents the temporal modulations up to the critical fusion frequency, the observers would find no problem. However, this view needs update.

Neurophysiological response measurements and psychophysical studies of visual adaptation effects to high-frequency stimuli suggest that the early visual system can respond to frequencies considerably higher than the critical fusion frequency [4][5]. Since the perceptual decision as to whether the stimulus alternation is visible or not is made through the entire processing of the visual system, the critical fusion frequency for fast stimuli is influenced not only by the initial temporal characteristics but also by the mid-level perceptual process. Specifically, the perceptual grouping process affects temporal integration/segregation by facilitating temporal grouping of the signals that are inferred to belong to the same event, while facilitating temporal segmentation of signals that are inferred to belong to different events [6].

For example, when two colors (e.g., red and green) alternate within a small area, they are more likely to be perceptually fused when presented against a dark background than against a mixed-color (yellow) background of equal luminance, even at the same temporal frequency [7]. This is presumably because the flicker on the dark background is more likely perceived as belonging to the identical object, and the grouping process facilitates color integration on the time axis within the object. Furthermore, this effect can be seen even when the flickering area is moving on the retina. This implies that color information is integrated along a specific motion trajectory (“trajectory integration”). Color signals belonging to different trajectories tend to be segregated even if they are temporally close to each other on the retina, which leads to the apparent increase in the critical fusion retinal frequency for moving stimuli than for flickering stimuli.

Trajectory integration is an important mechanism for the clear perception of moving images. The reason why the visual system integrates signals over time is partly a reflection of its physical temporal characteristics, but it also serves the function of removing the effects of noise and increasing sensitivity by taking a time average. It is the same principle as taking pictures of stars in the night by increasing the exposure time. However, simply extending the time of integration on the retinal coordinate will result in image blurring when the object under observation moves. The motion blur is produced by disagreement between the direction of integration and the direction of object motion. Trajectory integration improves sensitivity without inducing motion blur.

The temporal resolution reflects the result of complex and functional visual processing. Its properties cannot be predicted by a simple temporal low-pass filter.

In addition, the human eye is always moving. Eye movements can make visible high-frequency changes beyond the normal temporal resolution, as in the color-breaking problem with DLP.

The temporal characteristics of the human visual system are much more complex than described in the textbooks, leaving room for innovation in the future.

5 Material perception and display gamma value

The gamma value of the display should be properly adjusted to compensate for the non-linear response of the image-formation system, including the camera, to achieve a linear input-output relationship. Otherwise, the image will not be reproduced correctly. Nevertheless, it seems that we humans are not very sensitive to gamma value changes on the display. When directly comparing two images with different gamma values, one may be able to detect the difference. However, when only a single image is presented, we tend to perceive the image as right and natural despite a change in gamma value.

One reason for our insensitivity to gamma changes may be that the response of the visual system to input intensity is non-linear. Due to adaptation and spatial context effects, the same physical luminance does not look the same intensity. Another reason, which has not been widely recognized and I want to explain below, is related to a specific effect of gamma on human image interpretation.

The CG image rendering first specifies the shape of the object, the reflection property (material) of the object, and the lighting environment, and then simulates the behaviors of the lights present in the scene. The images that are physically projected onto the retina in everyday scenes are formed similarly, being produced by complex optical interaction of shape, material, and illumination. The task of the human visual system to perceive the world from the retinal image is to follow this image formation process in the reverse direction and estimate shape, material, and illumination (inverse optics or inverse rendering).

Let me consider the estimation of reflectance properties (material) from the image. In addition to the mean reflectance, which determines the brightness, and the spectral reflectance, which determines the color, there are complex reflection functions, described by bidirectional reflectance distribution function (BRDF) or bidirectional scattering-surface reflectance distribution function (BSSRDF), that produce material properties such as gloss and translucency. It is very difficult to infer these complex reflectance functions from a single image. We pointed out before that we humans may perceive material using image statistics correlating with the reflection function [8][9].

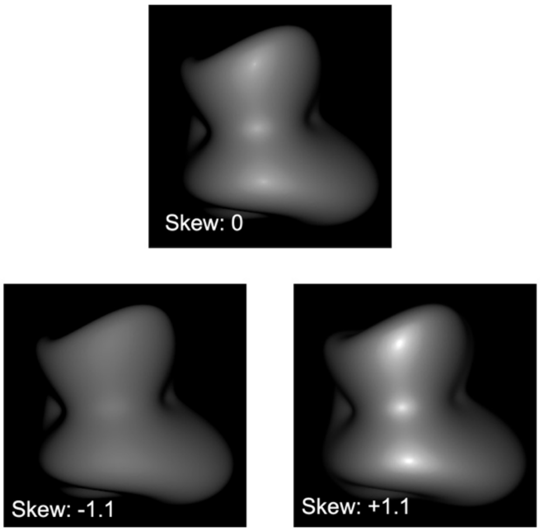

Fig. 1 Changing skewness of the luminance histogram affects perception of the material, but not that of the shape, of the object. Figure modified from Fig 9 in Ref [11].

For example, the skewness of the luminance histogram is related to the glossiness [8]. When the luminance histogram is manipulated so that it is more positively skewed (i.e., the skirt of the histogram widens on the high luminance side), the glossiness is enhanced (**Fig. 1**). Furthermore, when the color saturation is increased in addition to the skewness of the luminance histogram, the object is perceived as wet [10].

This operation of skewing the luminance histogram in the positive direction is the same as the operation of increasing the gamma value (expanding non-linearity) of the display. Therefore, manipulating the gamma affects the material perception. On the other hand, manipulating the skewness of the luminance histogram affect little the apparent shape of the visible object. What does this mean?

Object shape and material interact with each other to produce an image. To estimate both of them at the same time, the visual system may decompose images into luminance order information and luminance gradient magnitude information and estimate object shape from luminance order information and material from luminance gradient magnitude [11].

The information of luminance order includes the position and orientation of edges and contours. Changing the shape of the luminance histogram (including the display gamma manipulation) does not alter the luminance order. Human shape estimation uses image features that are robust to gamma changes.

On the other hand, as for materials, when specular reflection components are enhanced relative to diffuse reflection components in the surface BRDF, the shading luminance gradient becomes steeper in the image. This corresponds to an increase in the skewness of the luminance histogram, as well as an increase in display gamma. Therefore, material estimation uses the image features that are affected by gamma changes.

Display gamma manipulation mainly affects the apparent material of the image. There are many different materials in the world, and surface reflective properties vary greatly depending on the surface conditions (e.g., wet, dry). Even if the material looks different from the original one due to a gamma change, the human observers do not find problems in the image. I believe this is why we are not very sensitive to display gamma.

6 Concluding remarks: future direction

As we have discussed so far, visual science is a treasury of knowledge that is beneficial for display development. However, there is a wide range of knowledge in vision science. Many engineers may feel that it is impossible to learn all of them before developing their imaging technology. On the other hand, most vision researchers are specialized, and even if they know specific issues of their expertise, they cannot predict how the visual system as a whole will react to an arbitrary image.

To improve this situation and to link visual science and display development more seamlessly, I think it may be necessary to build a model that integrates a variety of findings of visual science. The model needs to be image-computable, being able to predict human perceptual responses to arbitrary images. The model can be thought of as a “digital twin” of the human visual system. As for the ocular system for retinal image formation, a project called "Isetbio" has developed a nice model [12]. For the modeling of the following neural information processing, an approach using artificial neural nets and machine learning has been proven to be useful [13].

Once we have a good image-computable model of the human visual system, we can incorporate it into the algorithm for computing image signals and optimize the parameters of the system including the observer [2]. It would also be possible to incorporate human models into adversarial training to automatically find algorithms that successfully fool humans.

With this goal in mind, I would like to continue my research on human vision.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP20H00603, JP20H05957.

References

[1] T. Kawabe, T. Fukiage, M. Sawayama & S. Nishida, “Deformation Lamps: a projection technique to make static objects perceptually dynamic,” ACM Trans. Appl. Percept., Vol. 13, No. 2, No. 10, pp. 1–17 (2016).

[2] T. Fukiage, T. Kawabe & S. Nishida, “Perceptually Based Adaptive Motion Retargeting to Animate Real Objects by Light Projection,” IEEE Trans. Vis. Comput. Graph., Vol. 25, No. 5, pp. 2061–2071 (2019).

[3] T. Fukiage, T. Kawabe & S. Nishida, “Hiding of phase-based stereo disparity for ghost-free viewing without glasses,” ACM Trans. Graphics, Vol. 36, No. 4, pp. 1–17 (2017).

[4] M. Gur & D.M. Snodderly. “A dissociation between brain activity and perception: chromatically opponent cortical neurons signal chromatic flicker that is not perceived,” Vol. 37, No. 4, pp. 377–382 (1997).

[5] E. Vul & D.I. MacLeod, “Contingent aftereffects distinguish conscious and preconscious color processing,” Nature Neuroscience, Vol. 9, pp. 873–874 (2006).

[6] S. Nishida, J. Watanabe, I. Kuriki & T. Tokimoto, “Human visual system integrates color signals along a motion trajectory,” Curr Biol, Vol. 17, No. 4, pp. 366–372. (2007)

[7] J. Watanabe & S. Nishida. “Veridical perception of moving colors by trajectory integration of input signals,” J. Vis., Vol. 7, No. 11, pp. 3:1-16 (2007).

[8] I. Motoyoshi, S. Nishida, L. Sharan & E.H. Adelson, “Image statistics and the perception of surface qualities,” Nature, Vol. 447(7141), pp. 206–209 (2007).

[9] S. Nishida, “Image statistics for material perception,” Current Opinion in Behavioral Sciences, Vol. 30, pp. 94–99 (2019).

[10] M. Sawayama, E.H. Adelson & S. Nishida, “Visual wetness perception based on image color statistics,” J. Vis., Vol. 17, No. 5, pp. 7:1–24 (2017).

[11] M. Sawayama & S. Nishida, “Material and shape perception based on two types of intensity gradient information,” PLoS Comp. Biol, Vol. 14, No. 4, e1006061 (2018).

[12] N.P. Cottaris, H. Jiang, X. Ding, B.A. Wandell, & D.H. Brainard, “A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine,” J Vis, Vol. 19, No. 4, pp. 8–27 (2019).

[13] N. Kriegeskorte & T. Golan, “Neural network models and deep learning,” Curr Biol, Vol. 29, No. 7, pp. R231–R236 (2019).