

1

Introduction

Camera imaging technology has evolved from a time-consuming, multi-step chemical analog process to that of a nearly instantaneous digital process with a plethora of image sharing possibilities. Once only a single-purpose device, a camera is now most commonly part of a multifunctional device, for example, a mobile phone. As digital single lens reflex (DSLR) cameras become more sophisticated and advanced, so also mobile imaging in products such as smartphones and tablet computers continues to surge forward in technological capability. In addition, advances in image processing allow for localized automatic enhancements that were not possible in the past. New feature algorithms and the advent of computational photography, for example, sophisticated noise reduction algorithms and post-capture depth processing, continue to flood the market. This necessitates an ever expanding list of fundamental image quality metrics in order to assess and compare the state of imaging systems. There are standards available that describe image quality measurement techniques, but few if any describe how to perform a complete characterization and benchmarking of cameras that consider combined aspects of image quality. This book aims to describe a methodology for doing this for both still and video imaging applications by providing (1) a discourse and discussions on image quality and its evaluation (including practical aspects of setting up a laboratory to do so) and (2) benchmarking approaches, considerations, and example data.

To be most useful and relevant, benchmarking metrics for image quality should provide consistent, reproducible, and perceptually correlated results. Furthermore, they should also be standardized in order to be meaningful to the international community. These needs have led to initiatives such as CPIQ (Camera Phone Image Quality), originally managed by the I3A (International Image Industry Association) but now run as part of standards development within the IEEE (Institute of Electrical and Electronics Engineers). The overall goal of this specific CPIQ work is to develop an image quality rating system that can be applied to camera phones and that describes the quality delivered in a better way than just a megapixel number. In order to accomplish this, metrics that are well-correlated with the subjective experience of image quality have been developed. Other imaging standards development includes the metrics by Working Group 18 of Technical Committee 42 of the International Organization for Standardization (ISO) and the International Telecommunication Union (ITU). These standards bodies have provided, and continue to develop, both objective and subjective image quality metrics. In this context, objective metrics are defined measurements for which the methodology and results are independent of human perception, while subjective metrics are defined measurements using human observers to quantify human response.

In following chapters, the science behind these metrics will be described in detail and provide groundwork for exemplary benchmarking approaches.

1.1 Image Content and Image Quality

Before delving into the specifics related to objective and subjective image quality camera benchmarking, exploration of the essence of photography provides justification, motivation, and inspiration for the task. As the initial purpose for photography was to generate a permanent reproduction of a moment in time (or a series of moments in time for motion imaging), an understanding of what constitutes the quality of objects in a scene will necessitate what to measure to determine the level of image quality of that permanent reproduction. The more a photograph or video represents the elements of a physical scene, the higher the possible attainment of perceived quality can become.

The efforts to create the first permanent photograph succeeded in the mid-1820s when Nicéphore Niépce captured an image of the view from his dormer window—a commonplace scene with buildings, a tree, and some sky. The image, produced by a heliographic technique, is difficult to interpret when observing the developed chemicals in the original state on a pewter plate (see Figure 1.1). In fact, the enhancement of this “raw” image, analogous to the image processing step in a digital image rendering, produces a scene with more recognizable content (see Figure 1.2). But, even though key elements are still discernible, the image is blurry, noisy, and monochrome. The minimal sharpness and graininess of the image prevent discernment of the actual textures in the scene, leaving the basic shapes and densities as cues for object recognition. Of note is the fact that the west and east facing walls of his home, seen on the sides of the image, are simultaneously illuminated by sunlight. This is related to the fact that the exposure was eight hours

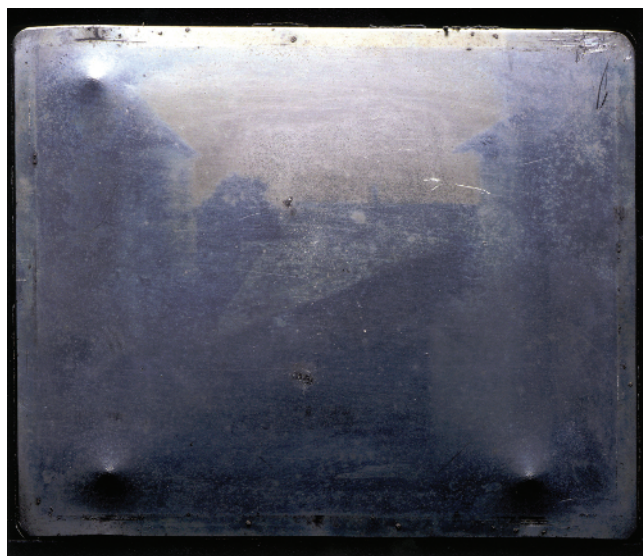


Figure 1.1 Image of first permanent photograph circa 1826 by N. Niépce on its original pewter plate.
Source: Courtesy of Gernsheim Collection, Harry Ransom Center, The University of Texas at Austin.

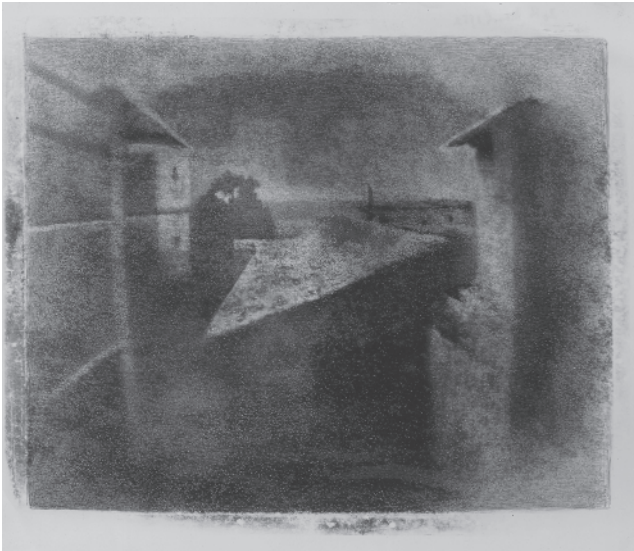


Figure 1.2 Enhanced version of first permanent photograph circa 1826 by N. Niépce. *Source:* Courtesy of Gernsheim Collection, Harry Ransom Center, The University of Texas at Austin.

in length, during which the sun's position moved across the sky and exposed opposing facades (Gernsheim and Gernsheim, 1969). Needless to say, the monochrome image is void of any chromatic information.

That we can recognize objects in the rustic, historic Niépce print is a comment on the fundamentals of perception. Simple visual cues can convey object information, lighting, and depth. For example, a series of abstract lines can be used to depict a viola as shown in Figure 1.3. However, the addition of color and shading increases the perceived realism of the musical instrument, as shown in the center image. A high quality photograph of a viola contains even more information, such as albedo and mesostructure of the object which constitute the fundamental elements of texture, as shown on the right. Imaging that aims for realism contains the fundamental, low level characteristics of color, shape, texture, depth, luminance range, and motion. Faithful reproduction of these physical properties results in an accurate, realistic image of scenes and objects. These properties will be described in general in the following sections and expanded upon in much greater detail in later chapters of the book, which define image quality attributes and their accompanying objective and subjective metrics.

1.1.1 Color

Color is the visual perception of the physical properties of an object when illuminated by light or when self-luminous. On a basic level, color can describe hues such as orange, blue, green, and yellow. We refer to objects such as yellow canaries, red apples, blue sky, and green leaves. These colors are examples of those within the visible wavelength spectrum of 380 nm to 720 nm for the human visual system (HVS). However, color is more complex than perception of primary hues: color includes the perception of lightness and brightness, which allows one to discriminate between red and light red (i.e., pink), for example, or to determine which side of a uniformly colored house is facing



Figure 1.3 Three renditions of a viola. Left: line sketch; middle: colored clip art (Papapishu, 2007); right: photograph. Each shows different aspects of object representation. Source: Papapishu, <https://openclipart.org/detail/4802/violin>. CC0 1.0

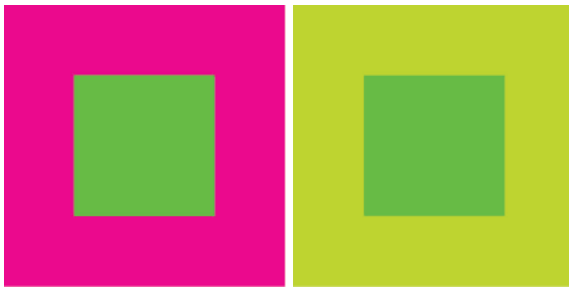


Figure 1.4 Example illustrating simultaneous contrast. The center squares are identical in hue, chroma, and lightness. However, they appear different when surrounded by backgrounds with different colors.

the sun based on the brightness of the walls. These are relative terms related to the contextualized perception of the physical properties of reflected, transmitted, or emitted light, including consideration of the most luminous object in the scene. Color perception is also impacted by the surrounding colors—even if two colors have the same hue, they can appear as different hues if surrounded by different colors. Figure 1.4 shows an example of this phenomenon called *simultaneous contrast*. Note in this example that the center squares are identical. However, the surrounding color changes the appearance of the squares such that they do not look like the same color despite the fact that they are measurably the same.

There are other aspects of the HVS that can influence our perception of color. Our ability to adapt to the color cast of our surroundings is very strong. This *chromatic adaptation* allows us to discount the color of the illumination and judge color in reference to the scene itself rather than absolute colorimetry. When we are outside during sunlight hours, we adapt to the bright daylight conditions. In a similar manner, we adapt to indoor conditions with artificial illumination and are still able to perceive differences in color. Perceptually, we can discern colors such as red, green, blue, and yellow under either condition. However, if we were to measure the spectral radiance of a physical object under two strongly varying illuminant conditions, the measurements would be

Figure 1.5 Example illustrating chromatic adaptation and differences between absolute and relative colorimetry. The fruit basket in the original photo clearly exhibits varying hues. A cyan bias is added to the original photo to generate the middle photo. With chromatic adaptation, this photo with the cyan cast will have perceptible hue differences as well, allowing the observer to note a yellowish hue to the bananas relative to the other fruit colors. However, the bottom photo illustrates that replacing the bananas in the original photo with the cyan-cast bananas (the identical physical color of the bananas in the middle cyan-cast photo) results in a noticeably different appearance. Here, the bananas have an appearance of an unripe green state because chromatic adaptation does not occur. *Source:* Adapted from Fairchild 2013.



substantially different. An example is presented in Fairchild (2013) in which a fruit basket that is well-balanced for daylight exhibits distinct hue differences among the fruit. This is illustrated in the top photo in Figure 1.5. Relative to other fruit in the basket, apples on the right look red, oranges look orange, bananas look yellow, and so on. A cyan cast can be added to the photo such that its overall appearance is distinctly different from the original photo. However, with some time to adapt to the new simulated illumination conditions as presented in the middle photo, chromatic adaptation should occur, after which the fruit will once again begin to exhibit expected relative color such as the bananas appearing to have a yellowish appearance and the apples on the right having a reddish appearance. If, however, the bananas (only) in the original photo are

replaced with those having the cyan cast, the chromatic adaptation does not take place; the bananas take on an unripe green appearance relative to the other fruit colors. So, too, the physical spectral reflectance is distinctly different for the bananas in the original and cyan-cast versions, though interpreted differently in the middle and bottom examples.

At times, due to the adaptive nature of the HVS, we can perceive color that is not physically present in a stimulus. A physiological example is part of our viewing experience every day, though we don't usually make note of the phenomenon. The signal of light detection in the eye travels to the brain via the optic nerve. This region is a *blind spot* in our vision because there are no light sensors present in the retina in this position. However, the HVS compensates for the two occlusions (one from each eye) and fills in the regions with signals similar to the surrounding areas such that the occlusions of the optic nerves are not normally noticed. This filling in phenomenon encompasses both color and texture. In fact, the HVS is even adaptable to the level of filling in blindspots with highly detailed patterns such as text (though experimental observers could not actually read the letters in the filled-in region) (Ramachandran and Gregory, 1991)! Therefore, it should not be surprising that there are conditions that can result in the HVS filling in information as the signal to the eye is processed even if a blindspot is not present. As such, there can be a perception of a color even when there is no physical stimulus of a hue. An example of such a phenomenon is the *watercolor illusion* in which the HVS detects a faint color filling in shapes which have an inner thin chromatic border of the perceived hue surrounded with an adjacent darker border of a different hue. The filled region's hue is lighter than the inner border, however. Figure 1.6 shows shapes with undulating borders, which typically instill stronger filling in than linear borders. As should be seen due to the illusion, the regions within the shapes have an apparent watercolor-like orange or green tint whereas the regions outside of the shapes do not have this faint hue. However, the inside of the shapes are not orange or green; all regions on either side of the undulating borders are physically the same and would have the same colorimetric values if measured, that is, the value of the white background of the page.

An object has many physical properties that contribute to its color, including its reflectance, transmittance or emittance, its angular dependency, and its translucency. Thus, quantifying color has complexity beyond characterizing the spectral nature of

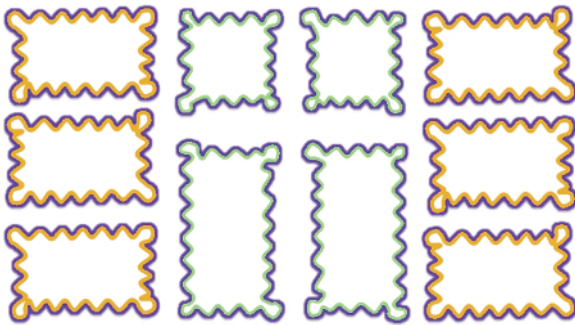


Figure 1.6 With a thin chromatic border bounded by a darker chromatic border, the internal region is perceived by the HVS to have a faint, light hue similar to the inner chromatic border even though the region has no hue other than the white background on the rest of the page. The regions within the shapes fill in with an orange or green tint due to the nature of the undulating borders and the hue of the inner border.

the color-defining element, such as a chromophore, dye, or pigment. Suppose we have a satin bed sheet and a broadcloth cotton bed sheet which are spectrally matching in hue, that is, having the same dye. However, we are able to discern a material difference because the satin sheet looks shiny and the broadcloth looks dull in nature. This difference in material appearance is because the satin has a woven mesostructure with very thin threads that generates a smooth, shiny surface when illuminated whereas the surface of the broadcloth is more diffuse due to thicker thread, lower thread count and a different weave, thus lacking the degree of shininess of satin. Yet, the color of the satin and broadcloth have matching color from a spectral standpoint. Another example of the complexity of color is the challenge of matching tooth color with a dental implant. Because teeth are translucent, the appearance of the whiteness is dependent on the lighting characteristics in the environment. Similar to placing a flashlight beam near the surface of marble, light can pass through a tooth as well as illuminate it. Thus, the challenge in matching a tooth appearance includes both a lightness and whiteness match as well as opaqueness. If a dental implant has a different opaqueness from the actual damaged tooth, there will be lighting environments in which the implant and tooth will not match even if the physical surface reflections of the white are identical.

Color measurements using colorimetry take into account the spectral properties of the illuminant, the spectral properties of the object, and the HVS. However, colorimetry has fundamental limitations when applied to the plethora of illuminants, objects, and people in the real world. In order to generate equations to estimate first-order color perception, data of (only) 17 color-normal observers were combined to generate the 1931 *standard observer* (Berns, 2000). That it was necessary to have more than one observer to make a standard observer is indicative of the inter-observer variability that exists in color perception. More recent works have confirmed that while this *observer metamerism* does exist, the 1931 standard observer remains a reasonable estimate of the typical color-normal observer (Alfvin and Fairchild, 1997; Shaw and Fairchild, 2002). In addition, inter-observer variability has been noted to be up to eight times greater than the differences inherent in the comparisons between the 1931 standard observer and five newer alternatives (Shaw and Fairchild, 2002). Thus, colorimetric quantification of colors incorporating the 1931 standard observer may predict color accuracy to a certain match level though an individual observer may not perceive the level as such. This becomes especially important considering the quality of colors in a scene that are captured by a camera and then observed on display or in printed material—the source of the colors of the scene, the display, and the printed material are composed of fundamentally differing spectral properties, but are assumed to have similar color for a high quality camera. In fact, color engineering could indeed have generated colors in a camera capturing system that match for the 1931 standard observer, but that matching approach does not guarantee that each individual observer will perceive a match or that the colorimetric match will provide the same impression of the original scene in the observer's mind.

Colorimetric equations are fundamental in quantifying the objective color image quality aspects of a camera. Measurements such as color accuracy, color uniformity, and color saturation metrics described later in the book utilize CIELAB colorimetric units to quantify color-related aspects of image quality. If, for example, the *color gamut* is wide, then more colors are reproducible in the image.

Quantifying the color performance, for example, color gamut, provides insight into an important facet of image quality of a camera system. However, as noted in previous examples, the appearance of color is more complex than the physical measurement of color alone, even when accounting for aspects of the HVS. Higher orders of color measurement include *color appearance models*, which account for the color surround and viewing conditions, among other complex aspects. Color appearance phenomena described in the examples above should point to the importance of understanding that sole objective measurements of color patches do not always correspond to the actual perception of the color in a photo. Challenges in measuring and benchmarking color will be discussed in more detail in further chapters.

1.1.2 Shape

A fundamental characteristic of object recognition in a scene is the identification of basic geometric structure. Biederman (1987) proposed a recognition-by-components theory in which objects are identified in a bottom-up approach where simple components are first assessed and then assembled into perception of a total object. These simple components were termed geometrical ions (or geons) with a total of 36 volumetric shapes identified, for example, cone, cylinder, horn, and lemon. Figure 1.7 has four examples showing how geons combine to form visually related, but functionally different, common objects. For example, in the center right a mug is depicted, whereas in the far right the same geons are combined to form a pail.

The vertices between neighboring geons are very important in distinguishing the overall object recognition: occlusions that overlap the vertices confuse recognition, whereas occlusions along geon segments can be filled in successfully (though this may require time to process perceptually). Biederman provides an example of the difference between these two scenarios (Biederman, 1987). Figure 1.8 contains an object with occluded vertices and a companion image in which only segments are occluded. The latter image on the right can be recognizable as the geons that comprise a flashlight whereas the former object is not readily discernible.

This bottom-up approach described above differs from Gestalt theory, which is fundamentally a top-down approach. “The whole is greater than the sum of the parts” is a generalization of the Gestalt concept by which perception starts with object recognition rather than an assimilation of parts. An example that bridges bottom-up and top-down theories is shown in Figure 1.9 (Carragher and Thurston, 1977). Top-down theorists point out that a Dalmatian emerges out of the scene upon study of the seemingly random

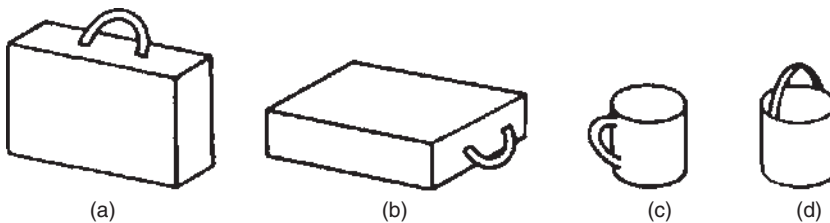


Figure 1.7 Examples showing how geons combine to form various objects. Far left: briefcase; center left: drawer; center right: mug; far right: pail. Source: Biederman 1987. Reproduced with permission of APA.

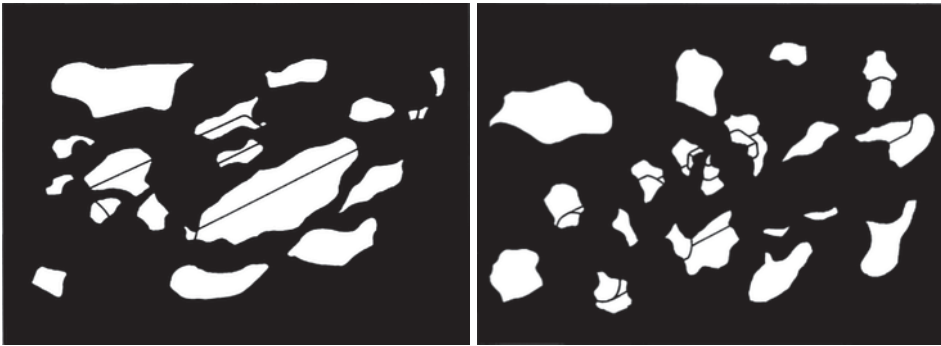


Figure 1.8 An example of an occluded object. Left: the vertices are occluded, making discernment of the object difficult. Right: only segments are occluded. In this right image, the object is more recognizable as a flashlight. *Source:* Biederman 1987. Reproduced with permission of APA.

Figure 1.9 An image associated with top-down processing in order to recognize the shape of a Dalmatian exploring the melting snow. *Source:* Republished with permission of John Wiley & Sons Inc, from *Optical Illusions and the Visual Arts*, Carragher and Thurston, Van Nostrand Reinhold Company, 7th printing, 1977; permission conveyed through Copyright Clearance Center, Inc.



collection of black blobs, while more recent research points to bottom-up processing for observers who found other objects in this scene such as an elephant or a jogger stretching out (van Tonder and Ejima, 2000). Regardless of the standpoint of bottom-up or top-down processing, shape is an important element of faithful scene reproduction.

Therefore, the spatially related aspects of an image will impact the perceived quality of the camera performance as pertaining to shape reproduction. Objective camera image quality metrics that are critical to shape quality include the spatial frequency response (SFR), resolution, bit depth, and geometric distortion. For example, a sharper image should increase the ability of the observer to see edges and, thus, shape and form in the image. Greater quality of shape and form, in turn, provides better camera image quality.



Figure 1.10 Influence of texture on appearance of fake versus real fruit. The fruits on the left in the top panoramic photo are all fake while the fruits on the right are real. Closer inspection of the pear surfaces can be seen in the bottom pair of images. The fake pear is on the left and the real pear is on the right. The texture appearance of the fake pear is composed of red paint drops.

1.1.3 Texture

Variations in apparent surface properties are abundant in both natural and synthetic physical objects. The HVS is adept at distinguishing these texture properties of objects. For example, in the field of mineralogy, an extensive vocabulary has been defined to describe the visual appearance of rock material (Adelson, 2001). These terms include words such as greasy, vitreous (glassy), dull, dendritic, granular, porous, scaly, and felted. While some of these terms such as greasy and scaly may conjure up specific visual differences, many of the mineralogists' terms refer to subtle changes in surface properties. This highlights the sophistication of the HVS as well as the importance of being able to generate realistic representations of objects in imaging systems. Appearance of material properties has been the focus of ongoing research and publications in the fields of perceptual psychology and computer graphics (Adelson, 2001; Landy, 2007; Motoyoshi *et al.*, 2007; Dorsey *et al.*, 2008; Rushmeier, 2008). Related to food appearance, there are fake products on the market that mimic real food. The top panoramic image in Figure 1.10 contains both fake and real fruits. Material properties that might provide clues as to which is which include texture and glossiness—attributes needing closer inspection. The bottom pair of images shows a crop of the fake pear surface on the left and the real pear surface on the right. In fact, the fake pear does have texture, but it is made with red paint drops whereas the real pear on the right has naturally occurring darker spots and even some surface scratches present in the lower right. As arranged in the panoramic photo at the top, the fake fruits are all on the left. This example shows that the appearance of material properties, for example, texture of fruits, influences the perception and interpretation of objects.

In photographic images, texture enhances object recognition. With changes in texture, an object can transform from appearing pitted and rough to appearing very smooth and



Figure 1.11 Left: the original image; right: after applying a sigma filter similar to one that would be used to reduce image noise (See Chapter 4 for more information on sigma filters.). Note the loss of texture in the hair, skin, and clothing, which lowers overall quality even though edges of the face, eyes, and teeth remain mostly intact.

shiny. Texture elements can also provide contextual information such as the direction of wind across a body of water. Many objects contain important texture elements such as foliage, hair, and clothing. Loss of texture in these elements can degrade overall image quality. As texture decreases, objects can begin to appear waxy and melted as well as becoming blurry. Figure 1.11 shows an example in which the original image on the left has been filtered on the right to simulate an image processing algorithm that reduces image noise (though in this particular example, the original image does not suffer from noise in order to accentuate the filtering result for demonstration). As can be seen, the filtering reduces the quality of the image because of blurring of the hair, skin, and clothing. Thus, objective image quality metrics that quantify texture reproduction are important for camera benchmarking.

1.1.4 Depth

Depth is an important aspect of relating to objects in the physical world. In a three-dimensional (3D) environment, an observer is able to distinguish objects in part by discerning the physical differences in depth. For example, an observer can tell which objects in a room may be within reach compared to objects that are in the distance due in part to binocular disparity of the left and right eyes. However, two-dimensional (2D) images are able to convey a sense of depth despite the lack of a physical third dimension. Several visual cues provide depth information in conventional pictorial images (Coren *et al.*, 2004):

- Interposition (object occlusion)
- Shading (variations in amount of light reflected from object surfaces)

- Aerial or atmospheric perspective (systematic differences in contrast and color when viewed from significant distances)
- Retinal and familiar size (size-distance relation based on angular subtense and previous knowledge of objects)
- Linear perspective (convergence of parallel lines)
- Texture gradient (systematic changes in size and shape as distance to viewer changes)
- Height in the plane (position relative to the horizon)

An additional visual cue for depth, specific to video imaging, is relative motion (motion parallax). When present, all of these visual cues are processed by the HVS in order to interpret the relationship between objects and illumination in the scene. However, because these are cues related to pictorial images, they are fundamentally monocular in nature. Thus, binocular aspects of the HVS, for example, convergence and binocular disparity, are not utilized to determine depth in these cases. In addition, the monocular function of lens accommodation for pictorial viewing is defined by physical distance to the picture, not by the various distances to objects that may be depicted in the scene. Thus, accommodation does not serve as a depth cue in the two-dimensional image scenario. However, realistic imaging is still able to convey depth and dimensionality with pictorial information void of 3D.

These visual cues for depth are dependent on camera image quality—images with elements such as sharp edges, high bit depth, and good color reproduction provide quality that is able to represent depth more fully even in a 2D scene. For example, an image that is blurry, low in bit depth, and monochrome has noticeably less perceptual depth to the objects in the scene compared to an image with high sharpness, sufficient bit depth, and color. Figure 1.12 contains a photograph pair demonstrating this comparison. The top image is monochrome, limited in bit depth, and noticeably blurry. In this image, the source of the surface modulations is non-obvious and the visual interpretation of the curvature and interposition has ambiguity. However, in the bottom version, the color and increased sharpness enable the viewer to better interpret the depth within the scene, including the structure of the sugar granules on the surface of the striped candy.

1.1.5 Luminance Range

Without illumination or self-luminance, scene content would not be discernible: light is a fundamental aspect of perception and imaging. Scene content contains objects that are illuminated or self-luminated by photons. The quantity of photons and surface reflectance or transmittance properties determine the luminance levels within a scene. For example, an object illuminated by candlelight will have a very small number of incident photons compared to the quantity when illuminated by sunlight. Color and surface properties determine the reflectance levels of the illuminated object. Shiny, metallic surfaces reflect a large percentage of incident light as do white, glossy objects. Dull, black objects and occlusions inhibit photon travel, resulting in low reflectance.

The HVS adapts to both light and dark conditions, expanding to an optimal range for a given environment (Fairchild, 2013). Yet, adaptation is not complete—this can be ascertained in one’s cognition of being in a moonlit environment versus a daylight environment. Thus, images should be able to represent both a form of absolute luminance and luminance range. If the camera’s exposure of a scene is not sufficient, the image will look too dark compared to an ideal representation of the scene or what the observer

Figure 1.12 Top: monochrome candy ribbons with low sharpness and bit depth, bottom: colorful candy ribbons with substantial sharpness. Note that the bottom image is more able to convey a sense of depth versus the top image.



recalls; the absolute luminance is not optimal. At worst, the image might be completely dark with indiscernible scene content. Conversely, if the camera's exposure of a scene is too high, the image will look too light at best and completely washed out at worst. For either case, the image quality can vary widely when observing an exposure series for a given scene.

Similarly, the lower the dynamic range of the rendered image, the more limited the image will be regarding representation of luminance range in the scene. As such, renderings with low dynamic range can lower the quality of scenes with high dynamic range. For example, glossy objects have high dynamic range when illuminated with direct light. Research has shown that rendering glossy objects with more dynamic range increases observer perception of glossiness (Phillips *et al.*, 2009). Thus, as an example, an image with lower dynamic range will have more limitations in representing the attribute of glossiness of an object compared to an image with higher dynamic range.

Figure 1.13 contains a tetrptych of images demonstrating variations in luminance levels and dynamic range. The first three images show an exposure series that shows how changes in the absolute luminance levels emphasize and reveal highlights and shadows in the scene: the underexposure by 2 f-stops of the camera allows one to see details in the shale gorge wall and sunlit trees in the background while the overexposure by 2 f-stops

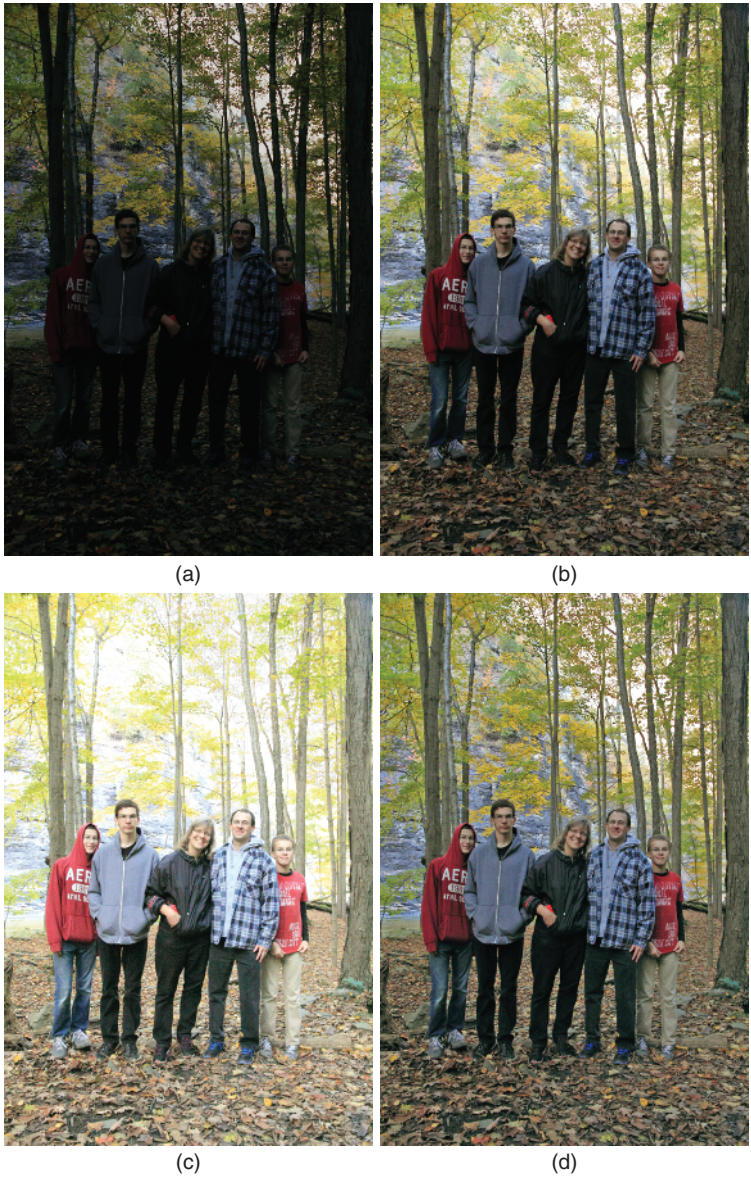


Figure 1.13 Variations in luminance levels and dynamic range for an example scene. (a) Underexposed by 2 f-stops. (b) Normal exposure. (c) Overexposed by 2 f-stops. (d) Normal exposure with localized tonemapping.

allows one to see details in the clothing on the models. The final image has localized tone mapping applied to the scene, which results in a rendition with optimized dynamic range in which more highlight and shadow details are apparent; this scene has optimized exposure and dynamic range, which in turn results in higher image quality.

1.1.6 Motion

Contributed by Hugh Denman

Motion within a scene is extremely informative for distinguishing and recognizing objects. Perception of motion allows us to determine critical aspects such as the velocity, that is, speed and direction, and the dimensionality of moving objects around us as well as depth in the scene. In this way, motion provides salient information to assess our environment. In fact, motion is of such critical importance that it is encoded by the HVS as a first-order, low level visual percept, similar to edge and texture perception. This contrasts with the naïve supposition that visual perception supplies a continuous stream of “images” of the scene, and that higher-level processes infer motion by comparing successive images (Sekuler *et al.*, 1990; Nakayama, 1985).

Broadly speaking, there are three sorts of stimuli which give rise to the perception of motion. The first may be termed “actual motion,” generated by moving elements in the scene or by the motion of the observer. The second is the well-known “apparent motion” effect. If a static stimulus is presented in a succession of spatial locations, with an interval of less than 100 ms between presentations, the stimulus will be perceived to move continuously, rather than being perceived to disappear and reappear in different locations (which, incidentally, is the percept if the interval increases above 100 ms). For example, a row of lights, each of which is briefly lit in succession in an otherwise dark scene, creates the impression of a single moving light if the delay between each light’s blink is less than 100 ms. This effect has long been exploited in visual entertainments, from flip-book animations to the kinetoscope and the cinema. While the term “persistence of vision” continues as a description of the apparent motion phenomenon, this dates to an early misconception of the eye as a sort of camera in which a retinal after-image is retained between stimuli (Anderson and Anderson, 1993). The more neutral term “beta movement” is preferred—Max Wertheimer coined this in the founding monograph of Gestalt psychology, *“Experimentelle Studien über das Sehen von Bewegung”* (Wertheimer, 1912).

The same monograph describes the third sort of motion stimulus, the “phi phenomenon,” in which the subject perceives motion without perceiving anything move. Consider a pair of stimuli, each depicting the same object (a small disc is often used), with a small spatial distance between the object positions. If these stimuli are presented in continuous alternation, with a very short switching interval (less than 30 ms), a flickering image of the object is perceived in both locations simultaneously—and a perception of motion between the object locations is also induced. This motion has no contour: the motion percept is not affected by the shape of the stimulus object (Steinman *et al.*, 2000). Because the motion percept is not associated with any moving object, Wertheimer termed this “pure” motion perception, and concluded that motion perception is “as primary as any other sensory phenomenon”.

It is now known that there is an area within the visual cortex, designated MT or V5, which encodes an explicit representation of perceived motion in terms of direction and speed. This area is also concerned with the somewhat related task of depth perception. The motion percepts arising here can be experienced “out-of-context” through various *motion aftereffect* visual illusions (Anstis, 2015). For example, if one stares at a waterfall for a few minutes and then looks away, a perception of upward motion is superimposed

on the scene—this is due to neuronal adaptation within area MT. There are accounts of patients with damage to this area who experience *akinetopsia*: the inability to perceive motion. Temporary disruptions to motion perception, arising from migraines or consumption of hallucinogens, are termed episodes of *cinematographic vision*.

Motion perception has been extensively studied using *random-dot cinematograms*. These consist of a series of images, each depicting a field of dots. Most of the dots appear in a different, random position in each image, but a subset of dots are made to move from image to image. A sample random-dot cinematogram is shown in Figure 1.14a and b. When presented statically, side-by-side, these two images should appear to contain entirely random dot fields. However, if presented in alternating superposition, one pair of dots will be seen to correspond (via motion) from frame to frame, while the others are appearing and disappearing. In Figure 1.14c and d, the pair of dots that have apparent motion have been highlighted in red.

The principal psychophysical parameter determined by these experiments is the maximum displacement at which motion can be detected, denoted $d_{\max}$. Dots that are displaced by more than this amount are not perceived as moving, but rather as disappearing and reappearing in a new location. That this sort of correspondence can be readily established in temporal succession, but not in spatial (i.e., side-by-side) presentation, is due to the specialized motion perception machinery of the HVS.

$d_{\max}$ increases with eccentricity (i.e., toward the periphery of vision), from about 9 minutes of arc at the center of the visual field, to about 90 minutes of arc at 10 degrees off-axis. Thus, motion perception is an attribute of vision whose performance improves off-axis, unlike most others such as color perception and acuity. $d_{\max}$ can also be increased by low-pass filtering the stimulus (for example, introducing a blur by squinting). This suggests that the presence of high spatial frequencies can prevent the perception of motion. Random-dot cinematograms have also been used to investigate *motion metamerism*: a pair of stimuli in which the dots follow distinct motion trajectories can induce indistinguishable motion percepts, if certain statistics of the motions are identical.

Stimuli giving rise to apparent motion effects, such as random-dot cinematograms as shown in Figure 1.14, can be ambiguous regarding the underlying, continuous motion paths. For example, a pair of dots displaced by the same distance from one image to the next could have traveled in parallel, or could have crossed paths *en route* to the new positions. This is shown in Figure 1.14e and f. Such ambiguities are resolved at a low level: there is no perception of ambiguous motion, nor a conscious choice of motion hypothesis. Thus, the machinery of motion perception consists not only of correspondence matching apparatus, but also apparently the imposition of constraining assumptions such as parsimony, inertia and rigidity of objects (Ramachandran and Anstis, 1986; Gepshtein and Kubovy, 2007).

Camera systems rely on the apparent motion effect to capture convincing video—the frame rate must be high enough to induce the motion percept. As mentioned above, this beta movement effect requires playback rates of about 10 frames per second, or higher. For example, cinema has traditionally used 24 frames per second (fps) for playback rate. In addition, for realistic motion presentation, the capture rate must match the intended playback rate. Thus, regarding benchmarking image quality, the camera frame rate capture and playback directly impact the visual quality.

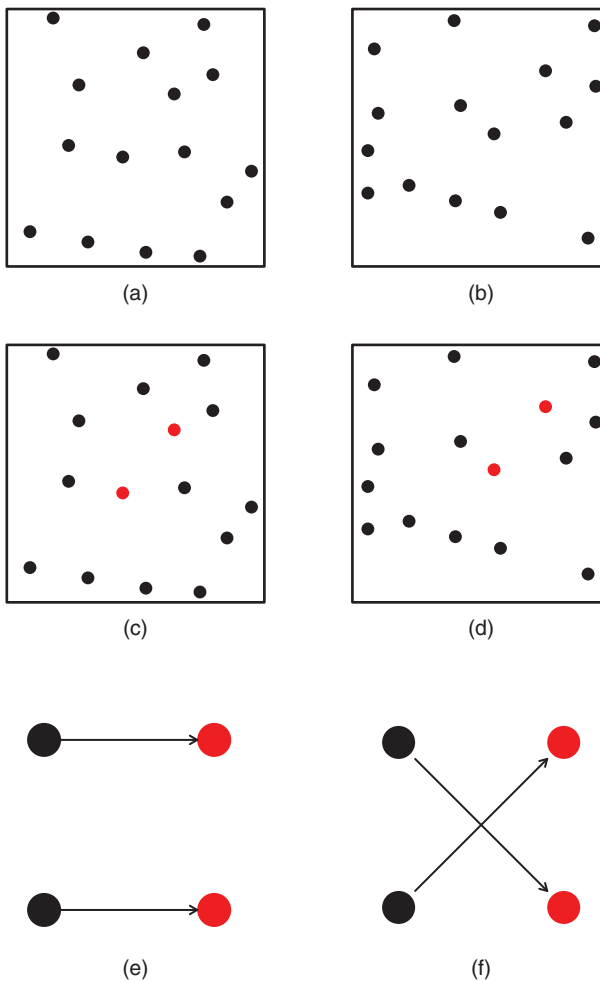


Figure 1.14 Random-dot cinematograms. (a) First frame of a two-frame cinematogram. (b) Second frame of a two-frame cinematogram. (c) The same frame shown in (a), with moving dots shown in red. (d) The same frame shown in (b), with moving dots shown in red. (e) A plausible motion hypothesis for a two-frame cinematogram in which the dots move from the positions in black to those in red. (f) Another plausible motion hypothesis for a two-frame cinematogram in which the dots move from the positions in black to those in red.

As well as the frame rate, the exposure time per frame (shutter speed) affects the perception of motion. Longer exposure times introduce *motion blur*, which increases the perceived smoothness of motion but reduces the visual detail in each frame. A lack of motion blur at lower frame rates (24–30 fps) can result in motion *judder*: jerky movement of objects in the scene. In cinema, the shutter speed is varied according to the motion content and directorial intent. Consumer cameras typically choose frame rate and frame exposure time automatically, to enable correct exposure according to the light level. This excludes the possibility of manipulating the quality of motion capture by manipulating these parameters.

It is clear that once the essential technical requirements for motion capture are met, the motion capture performance of a camera is highly dependent on the nature of the motion itself to be captured. Tests for the motion capture performance of cameras are not yet highly developed, and standardized motion test targets are only beginning to appear. However, several metrics for digital video quality are available and these can be utilized to assess the motion capture performance of a camera. These are discussed in later chapters.

1.2 Benchmarking

Now that we have explored these six key aspects of the essence of imaging—color, shape, texture, depth, luminance range, and motion—we can begin to explore the task of image quality benchmarking. Photographic technology has evolved immensely since Nicéphore Niépce captured the first permanent photograph with his camera obscura in the late 1820s. As mentioned at the beginning of this chapter, cameras have been primarily single-purpose devices over the past centuries: for capturing still images and/or video—though of varying complexity and capability. More recently, mobile phone cameras have evolved from low-resolution, low quality gadgets into fully-fledged photographic and videographic tools, dwarfing placement of traditional cameras in the marketplace. Because of this revolutionary development, the imaging industry has been revitalized regarding the necessity of being able to specify and characterize image quality in a reliable and consistent way, and in a way that also correlates with human vision.

The process of objective and subjective camera image quality benchmarking varies both in breadth and depth, depending on the intent of the benchmarking. A key component of benchmarking is determining what questions need to be answered—if one can envision the type of information needed from the process, then the steps to obtain the benchmarking will become clearer. For example, is the benchmarking intended to compare isolated components of the system such as the sensor or the optics? Then, objective metrics and specialized equipment for characterizing these components can be utilized. If, however, the intention follows the main topic of this book—that of camera system benchmarking—then the integrated behavior of the components needs to be incorporated into the analysis. Typically, this means quantification of key image quality aspects as listed below:

- Exposure and tone
- Dynamic range
- Color
- Shading
- Geometric distortion
- Stray light
- Sharpness and resolution
- Texture blur
- Noise
- Color fringing
- Image defects

The integrated measurement of these aspects of image quality provides a means of predicting how a consumer will perceive photographs from a particular camera. Subsequently, benchmarking is possible when comparing results from multiple cameras. Objective and subjective metrics used to quantify these attributes are described in detail in following chapters. Note that many of the attributes listed above are dependent on the spatial scale. For example, sharpness and resolution of an image will be perceived differently for a given photograph, depending on how close the image is viewed or how much magnification is applied to generate the size of the viewed image, whereas the quality of the exposure and tone of the image typically remains constant under these conditions. Thus, benchmarking should incorporate the use case of the still image or video clip in order to provide more meaningful and appropriate results.

In order to expand on this use case topic, let us return again to the early days of photography. The process to view early photography captured on glass plates was commonly one in which a photograph was generated by a contact print method—a print made by shining light through the glass negative placed in direct contact with the light sensitive emulsion layer of the paper. For this situation, no magnification of the image in the negative was applied to the viewed image: the size of the objects in the negative was the same as the size of the objects in the print. As glass plates transitioned to film made of flexible cellulose support, the print-making process commonly held to that of contact prints. However, as film evolved, film machinery configurations and standards led to the size of 35 mm film for motion pictures (Fullerton and Söderbergh-Widding, 2000; Dickson, 1933), which then became popularized by Leica for still photography. In order to make photographic prints from this film format, the film was no longer placed directly on the photosensitive paper, but was instead projected onto the paper from a distance by means of a lens. For this situation, the photograph became a magnified version of the image in the negative because the print size could be several times larger than the original image. As such, the image quality aspects of the photograph could differ significantly from what was directly measured in the film image. For example, to print a traditional 4R 4 × 6 inch (10 × 15 cm) print, the 35 mm negative is magnified approximately three times in height; to print a traditional 5R 5 × 7 inch (13 × 18 cm) print, the 35 mm negative is magnified approximately four times in height, and so on. Thus, small changes in spatially related image quality properties of the negative become increasingly more important as the source of the image becomes smaller and the size of the photographic output becomes larger, magnifying the aspects of these scale-dependent image quality attributes.

Considering how this aspect of magnification relates to the state of benchmarking digital camera image quality, suppose that one captures a digital image using a mobile phone. Early phones had cameras with digital sensor resolutions of 640 × 480 pixels, or 0.3 megapixels (MP), that is, not a lot of information compared to current camera phones with sensors that strongly exceed this resolution. Given that the resolution of the phone displays coincident with these VGA (Video Graphics Array) sensors were even *less*, the process of displaying a photo actually required a *downsampling* of the image. This reduction in pixel resolution in essence increased the perceived image quality of the scale-dependent attributes, such as sharpness and resolution, from what the sensor captured. Thus, it is not a surprise that people were disappointed with the quality of 4R 4 × 6 inch prints when they first tried printing photos from their early camera phones because the typical print assumed 1800 × 1200 pixels (2 MP) minimum resolution for baseline image quality and their cameras were only capturing 0.3 MP images.

Even with the current advancement of resolutions of camera phone displays and sensors, most often the screen resolution on phones is significantly lower than that of the sensor such that the display on a phone limits the consumer from viewing the native image quality of the camera. For example, a phone with a 4 MP display for a 16 MP camera would have to downsample the camera image height by approximately two times to show on the display for the typical use case of observing camera phone image quality. If, however, the consumer were to magnify the image by zooming in on the phone display, then the perceived quality for this use case would be closer to that of the native camera resolution. Another use case example is to observe the quality of an image on a computer display. Depending on the resolution, physical size, and viewing distance for a given monitor (among other conditions), the perceived image quality of a photo would vary. For example, 4k UHD (ultra high definition) monitors are 3840×2160 pixels (8 MP). For sensors with smaller resolution than 4k UHD, the image would have to be magnified. However, for sensors with a resolution of 8 MP or higher, the impact of resolution on image quality would be similar.

Thus, the current limitations of image quality are not really about the megapixel resolution of the sensor, at least for most use cases of cameras with 8 MP sensors or higher. Often, other performance factors of the camera system such as the pixel size, the full well capacity of the image sensor, optics, and the image signal processing (ISP) pipeline are the limiting factors above the sensor resolution. However, the use case(s) for the benchmarking will dictate how important the magnification (digital zoom) aspect is for comparing cameras.

Suppose you want a general benchmark comparison of how consumer cameras compare. One way to approach the task is to generate the image quality assessment for each camera given the type of scene content and application categories that are important to the consumer. The concept of *photospace*, based on the probability distribution of subject illumination level and subject-to-camera distance in photos taken by consumers, has been used to define the scene content categories that are important to include in development analysis related to benchmarking (Rice and Faulkner, 1983). For example, scene content such as a macro photo of a check for bank deposit, a photo of friends in a dimly lit bar, a typical indoor portrait, an indoor stadium sports event, and a daylight landscape photo are all common and important scenes for the typical consumer, representing various illumination levels and subject-to-camera distances. As such, these examples of photospace would provide salient scene categories to include in a comparative assessment.

Modes of viewing photos or videos of these scenes include applications such as viewing on the display of the source camera or camera phone, viewing on a tablet computer, viewing on a UHD monitor or television, or enlarging the photo to hang on a wall as artwork. A simple matrix example adapted from concepts by I3A CPIQ is shown in Figure 1.15 with image quality assessment of the various combinations of scene content and application use cases (Touchard, 2010). Various means can be used to populate the matrix such as an image quality scale value or simplified assessment such as symbols or colors conveying the benchmark assessment. From this type of assessment, a general benchmark comparison can be made between cameras for given combinations of scene content and application use cases.

Figure 1.15 This example benchmark matrix shows the image quality assessment of various scene content and application categories for a consumer camera. Note how the quality varies as these categories change. *Source:* Adapted from Touchard (2010).

	Phone Display	Tablet Display	4k TV	Enlarged Wall Art
Macro	★4	★4	★4	★2
Dim Bar	★3	★3	★2	★2
Indoor Portrait	★5	★5	★5	★4
Sports	★4	★4	★4	★3
Landscape	★5	★5	★5	★4

What one should notice in the example benchmark matrix in Figure 1.15 is that the uses cases include different aspects of scene content such as illumination level, distance from the camera to the subject, and motion of the subject as well as application of the image such as magnification of the viewed output. The key image quality aspects listed earlier in the chapter should be quantified for each use case in the matrix. Most commonly, this involves quantifying the behavior of a camera based on specific individual image quality metrics for these aspects of color, shape, texture, depth, luminance range, and motion.

Many image quality metrics exist for defining objective image quality. More recently, objective metrics have begun to expand into the realm of subjective evaluation, resulting in perceptually correlated image quality metrics. For example, ISO 15739, written by the Technical Committee 42 of the ISO, incorporates a noise metric extension that predicts the subjective impact of a noise pattern (ISO, 2013). In addition, the image quality metrics by the CPIQ working group of the IEEE Standards Association contain equations to predict the subjective quality loss to a photograph for a given metric value (IEEE, 2017). However, as technology continues to evolve, image quality attributes also continue to migrate, necessitating new and revised means of quantification. These challenges become continually important as new hardware and software aspects introduce more and more spatially localized characteristics into the images and video frames.

Because image quality is in essence a subjective matter, quantifying subjective image quality is just as important as quantifying objective image quality. Systematic science, as established in the field of psychophysics, can be used to measure and quantify what observers perceive about image quality. Chapter 5 will define and discuss this type of subjective evaluation as it relates to image quality metrics. Further discussion on subjective evaluation will continue in subsequent chapters. As noted above, a set of objective measurements can only address the image quality attributes being measured, which makes it possible that a benchmarking approach is not comprehensive. Thus, subjective evaluation should be incorporated into any comprehensive image quality benchmarking approach, either by ensuring that objective metrics contain perceptually correlated metrics or by including subjective image quality metrics themselves into the

benchmarking formula. The remainder of this book will spell out the reasoning and details behind this premise.

1.3 Book Content

The following section provides summaries of the remaining book content, providing the reader with a concise explanation of the aim of each chapter.

Chapter 2: Defining Image Quality

Chapter 2 will provide a broad overview of image quality as well as the necessary definitions of the key terms that will be used through the book. First, image quality itself will be defined; then we will define its attributes and how they are categorized. We will also define the difference between global and local attributes. A section will be devoted to defining the specifics and differences of objective versus subjective image quality assessment methods.

Chapter 3: Image Quality Attributes

In Chapter 3, we will describe image quality attributes in more detail. When attempting to quantify image quality using objective measurements, one usually divides the overall impression into several separate “nesses”—sharpness, graininess, colorfulness, and so on—examples of attributes of image quality (Engeldrum, 2000). Each of these attributes has their own distinct signature. Starting out from the categorization of local, global, or video-specific, this chapter will describe each of the attributes in detail, providing many example images and figures. The chapter will conclude with a discussion about measurable attributes versus unmeasurable artifacts for still images as well as video.

Chapter 4: The Camera

The fourth chapter will first describe the different hardware and software components that constitute a digital camera and its architecture. In particular, we will describe how digital camera components (the lens, the image sensor and the image signal processing (ISP)) all contribute to the performance and image quality of a camera. We will establish the connection between each component and the image quality attributes described in the previous chapter. Finally, for each component, we will detail the key parameters that influence image quality (e.g., aperture of the lens, etc.).

Chapter 5: Subjective Image Quality Assessment

Many psychophysical methods exist for quantifying subjective image quality with human observers. Chapter 5 will review key psychometric techniques, such as category scaling, forced-choice comparisons, acceptability ratings, and mean opinion score (MOS), and will emphasize the strengths and weaknesses of each methodology. The review will also explore the similarities and differences between still and video subjective evaluation techniques and how these are able to quantify important perceptual aspects of the human visual system’s assessment of image quality. Particular focus will be on the anchor scale method and how that can be used to quantify overall image

quality in just noticeable differences (JNDs) for still images in such a way that JNDs of various attributes can be combined to predict image quality of the camera.

Chapter 6: Objective Image Quality Assessment

Objective image quality metrics are by definition independent of human perception. Even so, by carefully choosing the methodology, it is possible to provide objective metrics that can be well-correlated with human vision. The content of Chapter 6 will provide an overview of existing metrics connected to the image quality attributes discussed in Chapter 3. We will describe in detail the “best” metrics for each of the attributes and also discuss pros and cons if there is more than one metric from which to choose for a given attribute. For instance, sharpness can be measured using resolution bars, sinusoidal Siemens stars, or slanted edges. Each of these methods will provide different results in some situations and it is important to understand the underlying reasons for the discrepancies. Moreover, practical issues, such as the choice of correct white point in color measurements, will be addressed in order to minimize the confusion which often arises because of the complexities.

Chapter 7: Perceptually Correlated Image Quality Metrics

In order for objective image quality metrics to be more meaningful to benchmarking, they need to be well-correlated with perception. Two approaches to accomplishing this are typically used, either through methods involving models of the human visual system, or by employing more empirical methods where some known aspect(s) of the human visual system can be taken into account, for example, correlations of adjacent pixels, and so on. Furthermore, some methods may be dependent on comparing the result to some known reference, while other methods may not. In Chapter 7, a large number of such methods will be discussed, including concepts such as mean square error (MSE) and peak signal to noise ratio (PSNR). We will also discuss methods to correlate the results of measurements on sharpness and noise to how these attributes are subjectively experienced. We will introduce the concept of contrast sensitivity functions (spatial and temporal), opponent color spaces, and so on, but also metrics mostly used in video quality assessment, such as the structural similarity index (SSIM) and similar. The importance of viewing conditions will also be stressed.

Chapter 8: Measurement Protocols—Building Up a Lab

When it comes to performing accurate and repeatable measurements, it is absolutely critical to establish and define the so-called protocols. The protocols provide a full description of the testing conditions that are required when performing image quality measurements. Chapter 8 will successively go over the protocols to be applied for objective and then subjective measurements. We will show how protocols are specific to each of the individual image quality attributes or parameters being measured. Discussion will include how protocols, such as those specifying lighting conditions, can vary as test equipment technology evolves.

Chapter 9: The Camera Benchmark Process

The first step to building a camera benchmark is to determine the key image quality attributes to be measured; then a method must be established to weight and combine

them to obtain a global scale so that one can benchmark all cameras against each other. In Chapter 9, we will show how a comprehensive camera benchmark should combine subjective and objective image quality assessment methodologies, and how some can substitute some others when correlation is established. We will describe the ideal benchmark and will show that, given the intrinsic subjectiveness of image quality, various approaches nearing the ideal might reach different conclusions. The chapter will also describe a number of existing camera benchmarking systems and will point to the ones that are the most advanced. Example benchmarking data will be shared for a collection of cameras, highlighting how various individual metrics can sway results. Finally, we will detail the possible evolution to move even closer to the ideal benchmark and highlight the technologies that remain to be developed to achieve this goal.

Chapter 10: Conclusion

The concluding chapter will restate the value and importance of a benchmarking approach that includes perceptually correlated image quality metrics. The section will also highlight future computational photography and hardware technologies that will be entering the mainstream consumer electronics market and how they impact the future of image quality metrics. Discussion will cover the challenges of benchmarking systems for the continually evolving camera imaging technology, image processing, and usage models.

Summary of this Chapter

- The more a photograph represents the elements of a physical scene, the higher the possible attainment of perceived quality can become.
- Key aspects of the essence of photography are color, shape, texture, depth, luminance range, and motion.
- Objective image quality evaluation involves making measurements, for which the results as well as methodology are independent of human perception.
- Subjective image quality evaluation is fundamentally a measurement quantifying human perception.
- Image quality is fundamentally a perceptual matter—it should include the perspective of an observer. Therefore, quantifying the subjective component is just as important as quantifying the objective component for the purpose of benchmarking image quality of cameras.
- To be most useful and relevant, benchmarking metrics for image quality should provide consistent, reproducible, and perceptually correlated results.
- Image quality is use case dependent: that which is deemed acceptable for one specific case may be unacceptable in other cases.
- The conditions under which a particular image or video is *captured* are important to define and understand when evaluating image quality. For example, camera performance under bright levels of illumination will almost certainly yield better image quality compared to capturing under dim levels of illumination.
- The conditions under which a particular image or video is *viewed* are important to define and understand when evaluating image quality. For example, viewing an image

on a mobile phone screen will almost certainly yield a different impression compared to a large format print of the image made on a high quality printer and hung on a wall.

- A set of objective measurements can only address the image quality attributes being measured, which makes it possible that a benchmarking approach is not comprehensive.
- Objective image quality metrics become more meaningful when the visual correlation is defined.
- Comprehensive benchmarks incorporate both objective and subjective image quality evaluation.

References

- Adelson, E. (2001) On seeing stuff: The perception of materials by humans and machines. *Proc. SPIE*, **4299**, 1–12.
- Alfvin, R.L. and Fairchild, M.D. (1997) Observer variability in metameric color matches using color reproduction media. *Color Res. Appl.*, **22**, 174–188.
- Anderson, J. and Anderson, B. (1993) The myth of persistence of vision revisited. *J. Film Video*, **45**, 3–12.
- Anstis, S. (2015) Seeing isn't believing: How motion illusions trick the visual system, and what they can teach us about how our eyes and brains evolved. *The Scientist*, **29** (6).
- Berns, R.S. (2000) *Principles of Color Technology*, John Wiley & Sons, Inc., New York, NY, USA, 3rd edn.
- Biederman, I. (1987) Recognition-by-components: A theory of human image understanding. *Psychol. Rev.*, **94** (2), 115–147.
- Carraher, R.G. and Thurston, J.B. (1977) *Optical Illusions and the Visual Arts*, Van Nostrand Reinhold Company, New York, NY, USA.
- Coren, S., Ward, L., and Enns, J. (2004) *Sensation and perception*, John Wiley & Sons, Inc., Hoboken, NJ, USA, sixth edn.
- Dickson, W.K.L. (1933) A brief history of the kinetograph, the kinetoscope and the kineto-phonograph. *J. Soc. Motion Pic. Eng.*, **21**, 435–455.
- Dorsey, J., Rushmeier, H., and Sillion, F. (2008) *Digital Modeling of Material Appearance*, Morgan Kaufmann Publishers, Burlington, MA, USA.
- Engel drum, P.G. (2000) *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Imcotek Press, Winchester, MA, USA.
- Fairchild, M.D. (2013) *Color Appearance Models*, John Wiley & Sons Ltd, Chichester, UK, 3rd edn.
- Fullerton, J. and Söderbergh-Widding, A. (eds) (2000) *Moving Images: From Edison to the Webcam*, vol 5, John Libbey & Company Pty Ltd, Sydney, Australia.
- Gepshtein, S. and Kubovy, M. (2007) The lawful perception of apparent motion. *J. Vision*, **7**, 1–15.
- Gernsheim, H. and Gernsheim, A. (1969) *The history of photography: from the camera obscura to the beginning of the modern era*, McGraw-Hill, New York, NY, USA.
- IEEE (2017) IEEE 1858-2016, IEEE Standard for Camera Phone Image Quality. IEEE.
- ISO (2013) ISO 15739:2013 Photography – Electronic Still Picture Imaging – Noise Measurements. ISO.
- Landy, M.S. (2007) Visual perception: A gloss on surface properties. *Nature*, **447**, 158–159.

- Motoyoshi, I., Nishida, S., Sharan, L., and Adelson, E.H. (2007) Image statistics and the perception of surface qualities. *Nature*, **447**, 206–209.
- Nakayama, K. (1985) Biological image motion processing: A review. *Vision. Res.*, **25**, 625–660.
- Papapishu (2007) <https://openclipart.org/detail/4802/violin>. (accessed 29 May 2017).
- Phillips, J.B., Ferwerda, J.A., and Luka, S. (2009) Effects of image dynamic range on apparent surface gloss, in *Proceedings of the IS&T 17th Color and Imaging Conference*, Society for Imaging Science and Technology, Albuquerque, New Mexico, USA, pp. 193–197.
- Ramachandran, V.S. and Anstis, S.M. (1986) The perception of apparent motion. *Scientific American*, **254** (6), 102–109.
- Ramachandran, V.S. and Gregory, R.L. (1991) Perceptual filling in of artificially induced scotomas in human vision. *Nature*, **350**, 669–702.
- Rice, T.M. and Faulkner, T.W. (1983) The use of photographic space in the development of the disc photographic system. *J. Appl. Photogr. Eng.*, **9**, 52–57.
- Rushmeier, H. (2008) The perception of simulated materials, in *ACM SIGGRAPH 2008 classes*, ACM, New York, NY, USA, SIGGRAPH '08, pp. 7:1–7:12.
- Sekuler, R., Anstis, S., Braddick, O.J., Brandt, T., Movshon, J.A., and Orban, G. (1990) The perception of motion., in *Visual perception: The neurophysiological foundations*, (eds L. Spillman and J. Werner), Academic Press, London, pp. 205–230.
- Shaw, M.Q. and Fairchild, M.D. (2002) Evaluating the 1931 CIE color-matching functions. *Color Res. Appl.*, **27**, 316–329.
- Steinman, R.M., Pizlo, Z., and Pizlo, F.J. (2000) Phi is not beta, and why Wertheimer's discovery launched the gestalt revolution. *Vision Res.*, **40**, 2257–2264.
- Touchard, N. (2010) I3A - Camera Phone Image Quality, in *Image Sensors Europe*, image-sensors.com.
- van Tonder, G. and Ejima, Y. (2000) Bottom-up clues in target finding: Why a Dalmatian may be mistaken for an elephant. *Perception*, **29**, 149–157.
- Wertheimer, M. (1912) Experimentelle studien über das sehen von bewegung. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, **61**, 161–265.