Faster and Higher

Optimizing Image Resolution

Barry Clark
barryclark@telenova.us
Rev. 04-15-11

Just when filmmakers thought it was safe to relax, imagining that the format wars were behind them, their idyll has been interrupted by the news that the prevailing standards for frame rate and resolution fall short of the requirements for optimum cinema and television presentations. Media watchers point to the inexorable march toward bigger, brighter, sharper, higher-contrast cinema and television screens, with many of these screens dedicated to 3D. And, according to imaging experts, for optimal presentation on screens such as these, films must originate at levels of resolution and (especially) frame rates that are beyond the technical capabilities of today's most advanced digital cinema cameras.

Temporal Resolution

One use of the term "temporal resolution" is to describe the frame rate of moving images[1]. The current 24 fps standard for frame rate was established some 85 years ago to standardize the projection speed for sound films. This standard—along with the 25 fps and 30 fps standards that were implemented for European (50 Hz region) and American (60 Hz region) television broadcasts—has long been the bane of filmmakers intent upon optimizing the audience's viewing experience. These frame rates, which amount to what Larry Thorpe of Canon has termed "severe temporal subsampling," are the source of the motion artifact known as flicker (or judder) and the source of the epithet "the flicks" that is aptly applied to movies. The effect, which is perceived not only as a flicker but as a loss of focus or definition, is detected when images of especially fast-moving subjects are displayed on a television or cinema screen[2]. The artifact becomes evident when the position of a subject in an individual frame is sufficiently distinct from its position in the previous frame that the eye-brain system is unable to fuse the two images into a semblance of uninterrupted motion. This "frame disparity" is most notable when a subject moves laterally (as opposed to diagonally) across the camera's field of view. The artifact is also especially apparent when objects are in sharp focus—a fact that provides one of the arguments in favor of large camera imagers and the shallow depth-of-field they afford. By employing shallow depth-of-field to achieve selective focus, filmmakers are able to isolate key subjects of interest and, at the same time, defocus fast-moving—and potentially distracting—subjects in the foreground and background of scenes.

In the real world, objects are free to move at any rate (up to the speed of light). But imaging systems—including the human eye-brain imaging system--are incapable of resolving images of subjects that move across our field of view at speeds that exceed a limit that is directly related to the frame rate (or "refresh rate") of the system. The average human eye is said to refresh at a rate of about 60 Hz (the equivalent of 60 fps if our eyes captured discreet frames), and exceptional eyes are said to refresh at rates that may be five times that fast. Individuals whose eyes refresh at very fast rates could wave their hands rapidly in front of their faces and not experience flicker or blur, while individuals whose eyes refresh at the average rate must wave their hands more slowly in order to hold them in focus. Because our 24 fps, 25 fps, and 30 fps acquisition and display systems have a flicker-fusion threshold that may be less than 50% of the threshold for the average viewer and as little as 10% of the threshold for the sharpest-eyed viewer, the rendering of fast-moving subjects in these systems differs significantly from the way that these viewers experience fast-moving subjects in real life.

---

[1] The term "temporal resolution," as applied to digital and film acquisition systems and to film projection systems, refers to the frame rate—i.e., the rate at which images are captured or displayed. In the case of digital display systems the term refers to the "refresh rate"-- the rate at which the systems write information to the screen. And in the case of the ultimate viewer, the term refers to the refresh rate of the eye-brain system—i.e., the rate at which information captured by the receptors in the viewer's eyes is discharged from these sites and downloaded to the brain. The term "temporal resolution" is one of those oxymorons on a par with the term "bandwidth," which is commonly used as a synonym for "bitrate." And the latter is itself an oxymoron, commonly used as a synonym for the space required to store a second of data. Add to these oxymorons the term "color depth," which is commonly used as a synonym for the number of bits assigned to color data processing, and you will need no further proof that the semantics of digital imaging is a jungle where the laws of common sense have no place.

[2] Productions for cinema release are normally captured at 24 fps using progressive-scan 1080p24 cameras, but (dependent upon the broadcast region) TV programming is normally captured either at 25 or 30 fps, using interlace-scan 1080i50 or 1080i60 cameras, or else it is captured at 50 or 60 fps, using progressive-scan720p50 or 720p60 cameras. Global HD broadcasting standards currently support frame rates of 25, 30, 50, and 60 fps via the 1080i50, 1080i60, 720p50, and 720p60 standards—with 1080p60 and 1080p50 standards yet to be implemented. While images captured in720p60 offer only 44% of the pixel count of comparable images in 1080p24, the 720p60 format, because of its higher frame rate (60 fps vs. 24 fps), produces images of fast-moving subjects that are generally judged to possess higher perceived resolution than images of the same subjects captured in 1080p24. A further penalty in perceived resolution is imposed by the capture of fast-moving subjects in the interlaced-scan formats, 1080i60 and 1080i50. For this reason ABC, Fox Sports, and ESPN in the U.S. opted early on to capture and transmit in 720p60 rather than 1080i60, and Sky Sports in the UK is reportedly weighing a switch from 1080i50 to 720p50.

As you can readily test by rapidly waving your hand close to your face and then waving it again at arm's length, the larger the display (i.e., the wider the angle-of-view) the more evident this disparity becomes. And, as can be confirmed by repeating the hand-waving test in a dimly-lit room, the brighter the display and the higher its contrast, the more evident the disparity becomes. Especially in a world of big, bright, high-contrast displays, the illusion of reality—the suspension of disbelief upon which film and TV depends—is compromised by the severe temporal subsampling inherent in our existing systems of image capture and display.

On TV and cinema screens the illusion of reality may be even more seriously compromised when images of fast-moving subjects are displayed in 3D, since the images from the right and left eyes may suffer not only from the previously-noted flicker effect but from the added "image-processing lag" caused by the disparity between the relative positions (or "parallax") of objects seen by the viewer's left and right eyes. For this reason, fast motion—and especially fast lateral motion across the frame, whether caused by the motion of the subject or the motion of the camera—is anathema in the world of big screen 3D productions.

Purists will insist that the flicker effect, along with the grain, the dust, the scratches, and the jump-and-weave introduced by the film camera and projector, contribute to the beloved "film look"--that elusive quality which is said to account for the rich emotional texture of the movies. But experts in optics note that the human visual system is remarkably plastic and, through experience, individuals may achieve levels of perception that once seemed far beyond their capabilities. In past decades movie audiences willing accepted the kind of crudely-painted studio backgrounds that would provoke laughter from contemporary moviegoers, while a generation of teenagers screamed in terror at horror-flick effects that they would now view as a joke. Yesterday's norm, the experts point out, in time becomes today's annoying artifact. And with moviegoers and TV viewers growing more media-fluent by the day, any inclination to underestimate or dismiss their visual sophistication could be a career-threatening miscalculation.

But if the effects of severe temporal subsampling are so egregious, why has the TV and cinema audience not risen up *en masse* to complain? Why do viewers willingly accept strobing images of waiters who rush past in the foreground of a restaurant scene; blurry images of gunmen who dart for cover on a city street; or images of diners who lose focus as they reach for a drink? Perhaps it is because, through long experience with artifacts like these, we have developed an internal "surge protector" that shields us from the effects of signal overloads. Such a system could, in effect, insert a "filter" into the datastream to the brain when we are confronted by rapidly-moving subjects, removing the filter only when the subject has slowed to a speed that permits it to be easily resolved. According to this view, audiences—whether consciously or not—may suffer from eye fatigue as a result of the effort they must expend in order to process such "extralimital" visual information. But, per this theory, once these audiences have experienced movies that render moving subjects with the fidelity that they have in real life, they will not willingly return to the energy-sapping world of the flicks[3].

Of course, many cineastes will protest that they go to the movies to *escape* from the real world, not to see the real world replicated on the screen. But just as the introduction of color and sound to the world of the movies did not preclude the production of both black & white and silent films, it is possible during production to blur, fog, filter, throw out of focus, or otherwise reduce the resolution of moving images if an impressionistic effect is desired. And flicker, dust, grain, scan lines, random noise, and other such artifacts can easily be introduced in post. But for those filmmakers who wish to abolish the proscenium arch that separates the playgoers from the play, the minimization of resolution-robbing artifacts is not just an obsession, it is a mission.

Spatial Resolution

The term "spatial resolution" commonly refers to the product of the vertical and the horizontal scan lines required to display an image. The spatial resolution of a 1080-line image amounts to 2.074 megapixels (1920 multiplied by 1080 lines), while the spatial resolution of a 720-line image is 0.922 MP (1280x720 lines). And the key standard-setting bodies (ITU in Europe, SMPTE in the U.S.) have now published specifications for 2K, 4K, and 8K systems, all with significantly higher temporal resolution than the current 1080 and 720-line broadcast standards. In the case of 8K the spatial resolution is 33.178 MP (7680x4320 lines), a figure that is 16 times the

[3]The trainability of the human eye can be demonstrated by the ability of most television viewers to adjust to the 30-frame and 25-frame flicker of 1080p60 and 1080p50 images, artifacts that can be evident to the untrained eye even in static images that have especially high levels of contrast and brightness. The worst effects of the 25-frame flicker are offset by 100 Hz displays in Europe, and the 24-frame flicker in theatrical films is offset to a degree by the standard practice of double or (in the case of 3D) triple "flashing" the projected frames—a technique that amounts to displaying each individual frame two or three times in quick succession. Some would claim, however, that the requirement made of the audience to subliminally fuse image information that is manipulated in this way may ultimately result in viewer fatigue. And they would say that--all other factors being equal--the viewer who can experience motion as it is experienced in real life and whose eyes are not required to perform such unnecessary visual work—whether at home or in the cinema--will find movies easier and more enjoyable to watch.

spatial resolution of a 1080-line HD system. Tests have shown that television viewers, watching screens with an aspect ratio of 1.78:1 from a viewing distance of three picture-heights or less, can detect the improvement in resolution that 4K affords relative to images displayed in the current 1080-line HD broadcast format. And similar tests have shown that cinema audiences, watching images from a distance of one picture-height or less, can easily detect the improvement that 8K affords relative to the current 2K digital cinema standard. But, as with temporal resolution, viewers with exceptionally acute vision are said to be able to discern levels of resolution that significantly exceed these limits[4].

While subsampling in the spatial domain may be less annoying to viewers than subsampling in the temporal domain, the inability of our television and cinema systems to display images that, from an average viewing distance (and with all other viewing and display variables equal) have a spatial resolution that is at or near the limits of the resolving power of our eyes, helps to maintain their "artifactual" nature—i.e., their inability to replicate the way we see the real world[5]. And, it is argued by the advocates of the optimum media experience, this limitation compromises the ability of filmmakers to persuade audiences to suspend their disbelief, to fully immerse themselves in the virtual worlds they have carefully created in their films.

Perceived Resolution

While the human visual system is only imperfectly understood, empirical tests suggest the way that variables such as frame rate and pixel count influence our perception of the overall "perceived" or "dynamic" resolution of a moving image. Such tests imply that the effects of temporal and spatial resolution, while not strictly independent of each other, may be considered to be multiplicative, with temporal resolution--in the particular case of fast-moving subjects--carrying more relative weight than spatial resolution in the determination of perceived resolution. This observation would suggest a relationship between the variables that is of the form $R_p \propto (\alpha R_t x \beta R_s)$, where R_p is an index of perceived resolution; R_t and R_s are indices of temporal and spatial resolution; and α and β are weighting factors, with $\alpha/\beta>1$ for images of fast-moving subjects[6].

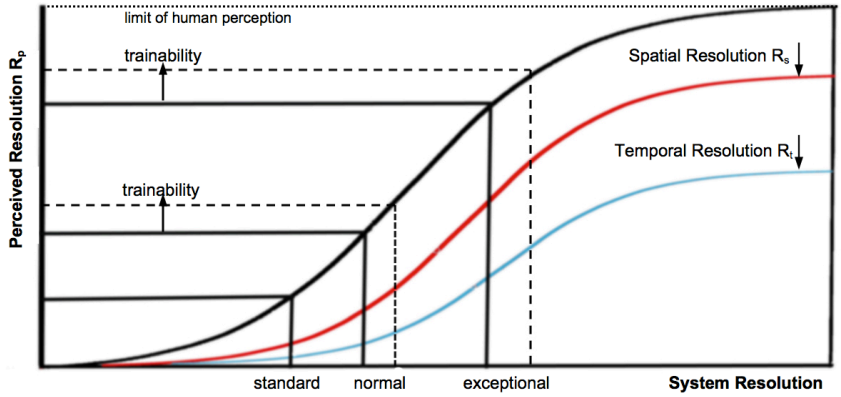
---

[4]The emergence of cinema image resolutions that are significantly beyond the limit of human perception will allow viewers to shift their attention to particular sectors of the screen without detecting annoying artifacts or blurred images. That is, viewers of large, high-resolution TV and cinema displays will be able to exercise the power of scrutiny or "close vision" in order to scan the entire area of the display instead of simply practicing the "single-gulp" style of viewing that is the norm with today's smaller, lower-res displays. Such a viewing method would have important practical and creative implications for filmmakers who have become accustomed to assuming that the viewers' eyes are fixed upon the center of TV and cinema displays.

[5]Note, in particular, the contribution of contrast to perceived resolution. A variable known as the contrast sensitivity function (CSF) provides a measure of how much contrast a viewer needs in order to distinguish a specific level of spatial resolution in an image. Though very fine detail may be present in an image, viewers may not be able to perceive all of this detail if the contrast level is too low. In this regard, a higher-contrast display with lower spatial resolution may display images that have higher perceived resolution than a lower-contrast display with higher spatial resolution. The contribution of contrast and other variables to perceived resolution is discussed in the "Additional Variables" section below.

[6]In the case of stationary and slow-moving images, the variable M--noted under "Additional Variables" below--is low, and perceived resolution is consequently high. But it's worth recalling that we don't call them the "movies" for nothing. Fast motion (even the motion of an on-camera host who ambles across the TV screen) is commonplace in the programming we watch. If, for such fast-moving subjects, we assume a conservative weighting ratio of $\alpha/\beta=1.5$, then (by the approximation suggested above, and with all other variables being equal) the R_p for the image captured and displayed in 720p60 would amount to 0.922 MP x 60 fps x 1.5 = 82.98 MP-frames/sec, while the R_p for the same image captured and displayed in 1080p24 would amount to 2.074 MP x 24 fps = 49.06 MP-frames/sec. That is, with all other variables being equal, the perceived resolution of the 1080p24 image amounts to only about 60% of that of the 720p60 image--a result that conforms to empirical observations. It should, however, be noted that the formula $R_p \propto (\alpha R_t x \beta R_s)$ is, at best, a rough approximation. Larry Thorpe points out that William Glenn has shown that the eye-brain system differs in its response to the temporal resolution of the luminance and chrominance components of a signal. In addition, other researchers have shown that human vision is more sensitive to the spatial resolution of the luminance component than it is to the spatial resolution of the chrominance component of a signal. Taking these observations into account, the temporal resolution term (R_t) in the formula above should, technically, comprise two variables, one for luminance (R_{tl}) and another for chrominance (R_{tc}). Similarly, the spatial resolution term (R_s) in the formula above should comprise two variables, one for luminance (R_{sl}) and another for chrominance (R_{sc})--with each of these variables assigned a weighting factor that reflects its relative contribution to the overall perceived resolution of the image (R_p).

Based on the above, for a viewing system of a fixed size and for a viewer at a fixed distance from a television or cinema display, the relationship between the temporal, spatial, and perceived resolution of the image of a fast-moving subject may be represented by a curve like the one below.



In the diagram above, the perceived resolution of the system is represented by the y-axis and the temporal and spatial resolution of the system is represented by the x-axis. The resolution value labeled "standard" represents the current television and cinema standards for spatial and temporal resolution; the value labeled "normal" represents the spatial and temporal resolution of the average human eye; and the value labeled "exceptional" represents the spatial and temporal resolution of an exceptional human eye. The dotted lines are intended to suggest the limits of trainability of normal and exceptional eyes. At the right side of the plot, where the resolution curves plateau, any further increase in the temporal or spatial resolution of the system will produce no further increase in the viewer's perceived resolution. That is, for a specific display size and viewing distance, the perceived resolution of the image of the fast-moving subject will have reached the limits of human perception.

Additional Variables

The perceived resolution of moving images on a television or cinema display depends upon a host of variables in addition to the temporal and spatial resolution of the images. Prominent among these are the accuracy of focus of the subject in the scene (F); the size (S)[7] of the imager in the camera; the brightness of the display (B); the brightness of the subject in the scene (B_1); the contrast level of the display (X); the contrast level of the subject (X_1); the rate of lateral motion (M) of the subject across the viewer's field of view; the chroma level of the display (C); the chroma level of the subject (C_1); the width of the display (W)[8] relative to a fixed viewing distance; the ratio of the size of the subject in the scene to the width of the display (W_1)); the presence of depth (3D)—contributed by a 3D display; the degree of parallax (P) (or separation) between the R and L "eyes" in a 3D display; the detail (or "sharpness") level of the display (D)[9]; the detail (or "sharpness") level of the subject (D_1); the noise level of the display (N); the noise level of the subject (N_1); the chromatic aberration contributed by the projector lens (CA_p); the chromatic aberration contributed by the camera lens (CA_c); the level of chroma compression (CC) employed in the camera; the employment of interlace scanning in image acquisition or display (I); the concatenation of resolution losses contributed by the analog components of the image processing path, from the lens of the camera to the lens of the projector (quantified by the modulation transfer function (MTF); and, lastly, the experience level of the viewer (E).

---

[7]Larger imagers (sensors) have larger photosites (pixels) than smaller imagers with the same number of photosites or pixels. The result is improved dynamic range, lower noise, and reduced crosstalk between adjacent photosites—all of which contributes to an increase in the perceived resolution of images from larger imagers.

[8]Note that, for a given viewing position, a wider display not only reveals more picture detail (or lack thereof) but, as noted above, viewers watching a wider display are more inclined to explore the screen instead of confining their gaze to the center of the screen. While the rods in our eyes are less sensitive to color than the cones in our eyes, they are more sensitive to motion at the periphery of our field of vision (an aid in detecting predators at the edge of our vision when we were hunting and gathering). As a consequence, on a wide display, or when viewing a smaller display from a close distance, the rods in our eyes are more likely to be exposed to the kind of motion that can result in the detection of annoying artifacts.

[9]Detail level, or "sharpness," is often confused with spatial resolution. But "sharpness," as colloquially used, is not a measure of the pixel count of an image but is instead a measure of the degree of edge-definition or detail enhancement that is applied to the image during post-processing.

Not all of the variables noted above deserve equal weight in the determination of perceived resolution, and not all of the variables are strictly independent of the rest. As an example, for a fixed bandwidth an increase in the brightness (or luma) of a signal is achieved only at the expense of the chroma component of the signal. And an increase in a display's level of image detail (or "sharpness") may be achieved at the expense of an increase in the level of noise in the image. But all of the variables noted here--among many others--should be taken into account when attempting to optimize the perceived resolution of a moving image.

While perceived resolution has been demonstrated to be directly proportional to the temporal and spatial resolution of a moving image, empirical observations suggest that it is inversely proportional to most of the variables listed above. Exceptions include the variable representing the accuracy of focus (F) of the subject and the presence of depth (S) in the display[10]. That is to say, perceived resolution is diminished by increasing the level of any of the variables in the denominator of the equation below; and, conversely, perceived resolution is enhanced by decreasing the level of any of the variables in the denominator of that equation. As an example, a viewer watching an old sitcom may experience an improvement in the perceived resolution of the program if it is viewed on a smaller display and the brightness, chroma, contrast, "sharpness," and noise in the image are all dialed down. And, it goes without saying, the perceived resolution of the image will be enhanced if the viewer lacks experience with the medium[11].

---

[10]The formula below suggests the relationship that appears to prevail between the most prominent factors that affect the perceived resolution of a moving image, viewed under fixed conditions and at a fixed distance from a display. The theoretical basis for the relationship between these variables remains to be established, and the empirical testing has yet to be carried out that would establish the relationship between the variables and assign appropriate weighting factors to each. But, in the absence of the necessary research, it may be a safe guess to say that the formula for perceived resolution would look something like this:

$$R_p \propto \left[\frac{f(\alpha R_t \cdot \beta R_s, F, S, 3D)}{f(B, B_1, X, X_1, M, C, C_1, W, W_1, P, D, D_1, N, N_1, CA_p, CA_c, CC, I, MTF, E} \right]$$

R_p = perceived resolution of the image
R_t = temporal resolution of the image
R_s = spatial resolution of the image
α = weighting factor for temporal resolution
β = weighting factor for spatial resolution
f = an indeterminate function of the variables listed within the brackets
F = accuracy of focus of the subject
S = size of the imager (sensor)
3D = presence of stereographic depth (3D)
B = brightness of the display
B_1= brightness of the subject
X = contrast level of the display
X_1= contrast level of the subject
M = rate of lateral motion of the subject across the field of view
C = chroma level of the display
C_1= chroma level of the subject
W = width of the display
W_1= ratio of the size of the subject to the width of the display
P = parallax, or distance between the image in the right and left "eyes" of a 3D display
D = detail level (or "sharpness") of the display
D_1 = detail level (or "sharpness") of the subject
N = noise level of the display
N_1= noise level of the subject
CA_p = chromatic aberration in the projector lens
CA_c = chromatic aberration in the camera lens
CS = chroma compression (subsampling) applied in the camera
I = employment of interlace scanning in acquisition and/or display
MTF = concatenation of resolution errors (or MTF penalties) in the imaging chain
E = experience level of the viewer

---

[11]Perceived resolution and "image quality" are not the same thing. "Quality" standards vary widely among social strata and cultures and, when asked to judge image quality, many viewers opt for high chroma values, high brightness and contrast levels, and high levels of detail enhancement or "sharpness"--all of which work to diminish what experienced viewers would call optimum resolution. Hence the "vivid" settings which are often set as the default picture mode on consumer TV sets.

Conclusion

With 3D promising to become commonplace and the brightness[12], contrast, and size of television and cinema displays increasing along with the experience level of the viewers, the perceived resolution of our entertainment media is demonstrably in decline. Cognizant of this, the cinema and television industries have begun to take the first small steps toward the implementation of a spatial resolution standard of 8K for cinemas, a 4K spatial resolution standard for homes, and a temporal resolution standard of 60 fps for both cinemas and homes. The realization of these goals is daunting to say the least and will depend upon, in addition to social and economic factors, significant improvements in the optics of acquisition and projection, as well as the increased availability of affordable solutions for data transmission and storage. But imaging technology continues to progress and the cost of bandwidth and storage has, for years, been in free-fall. While patience is required, these trends auger well for those who dream of a new level of immersion, credibility, and emotional impact from the media that mean so much to our lives.

Sources

Larry Thorpe, National Marketing Executive in the Broadcast & Communications Division of Canon U.S.A. and Steve Mahrer, Director of Engineering in the Business Development Group of Panasonic Broadcast & Television Systems Company U.S.A. kindly reviewed this document and pointed out needed corrections and clarifications. For further information, see the sizable body of work produced by William Glenn, Director of the Imaging Technology Center at Florida Atlantic University and his wife Karen. In addition, Larry Thorpe's informative white papers on image resolution and HDTV lens design contain a wealth of information. They can be accessed here:

http://www.cinematography.net/Files/Panavision/Sony/24PTechnicalSeminar2.pdf

and here:

http://www.abelcine.com/articles/index.php?option=com_content&task=view&id=43&Itemid=45

Another useful reference, particularly for its observations on the limits of human perception, is the 2008 NHK white paper on UHDTV resolution, available here:

http://www.smpte.org/journal/wp-content/uploads/2008/03/2008-04-uhdtv.pdf

The following paper discusses the contribution of 3D to perceived image quality:

http://www.ijsselsteijn.nl/papers/SPIEOpticsEast05.pdf

Feedback Is Invited

Comments and queries regarding the above may be directed to barryclark@telenova.us.

---

[12]Though current cinema projectors aim to deliver 16 ft-lamberts to the screen, luminance values of 60-70 ft-lamberts are now common for new TVs. 3D robs screens of luminance and, to compete with the bigger, brighter TV screens, the big theater chains may move toward screens that are up to 100' in width. To fill these screens with bright, high-resolution 3D images, new digital projection technologies will be required.