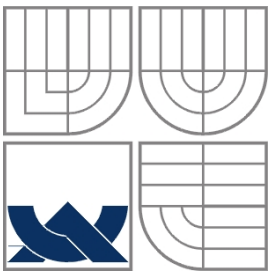
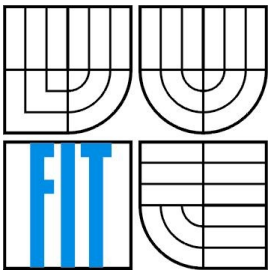


BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ



FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS
AND MULTIMEDIA

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

IMAGE PROCESSING FOR IMPROVED PERCEPTION AND INTERACTION

ZPRACOVÁNÍ OBRAZU PRO LEPŠÍ VJEM A INTERAKCI

DISSERTATION THESIS
DISERTAČNÍ PRÁCE

AUTHOR
AUTOR PRÁCE

Ing MICHAL SEEMAN

SUPERVISOR
VEDOUCÍ PRÁCE

Doc. Dr. Ing. PAVEL ZEMČÍK

BRNO 2012

Abstract

Image reproduction ought to provide subjective sensation possibly closest to the one, where the original image is observed. Digital image reproduction involves image capture, image processing and rendering. Several techniques in this process are not ideal. This work proposes improvement of speed and accuracy of some state-of-the-art methods.

Abstrakt

Reprodukce obrazu má zprostředkovat vjem co nejvíce podobný tomu, když pozorujeme původní obraz. Digitální reprodukce obrazu zahrnuje snímání, zpracování a vykreslení. Mnohé postupy v tomto procesu nejsou dokonalé. Tato práce předkládá zlepšení v rychlosti a přesnosti několika ze současných metod.

Keywords

visual perception, image processing, optimization, acceleration

Klíčová slova

vizuální vjem, zpracování obrazu, optimalizace, akcelerace

Citation

Michal Seeman: Image Processing for Improved Perception and Interaction, Dissertation thesis, Brno, FIT BUT, 2012

Image Processing
for Improved Perception and Interaction

Affirmation

I affirm, that this dissertation thesis was written only by myself,
under supervision of Doc. Dr. Ing. Pavel Zemčík.
A have been given another information by Ralph Nelson, Ph.D. and Dr. Ralph Pridmore
I have properly cited all sources.

.....
Michal Seeman
2012-06-29

Acknowledgments

I would like to thank my supervisor, Doc. Pavel Zemčík for indispensable help and support.
I would also like to thank to Ralph Nelson, Ph.D. for helping me with the retina response publications and Dr. Ralph Pridmore for providing his measurement chart.

Table of Contents

1 Introduction

Image reproduction is one of most common, most useful and most frequent tasks in human history. Technically, it means construction of a planar model aimed to evoke similar or nearly same sensation as with the observing of a real scene.

Digital image reproduction involves mainly image capture and image rendering. Between these two techniques, the data are digitally processed.

Meaning of the image processing might seem to be insignificant. In fact, if the image had been captured by an ideal camera and rendered via an ideal display device, no data processing would be necessary for perfect reproduction. Unfortunately, the available devices are certainly not ideal.

The scanning devices suffer of geometry distortion, luminance non-linearity and limited contrast (dynamic range). Although all of these imperfections has been overcome, in some of the cases it is at certain price. High dynamic range can be captured by multi-exposure, which does not allow for taking photographs of non static objects or capturing of motion pictures. Geometry correction can be measured and corrected, but the algorithms are rather slow for real-time processing.

Commonly used LCD display devices have pixel matrix fixed by construction so they do not suffer from geometry distortion. But the pixel density is still low, the matrix is visible and causes disturbing artifacts. Either the highest displayable contrast is still limiting. Despite of the marketing claims, most of the common displays are not capable of rendering much higher contrast than 1:1000. The ordinary practice is to scale range of the digital image to fit the display range. The procedure is so frequent, that many users do not consider it as an image processing at all. Yet the image is certainly changed. And as will be explained further in this work, substantial change in the contrast causes noticeable change in the color perception.

In some cases, the user demands to alter the image instead of perfect reproduction. The operation is not easy, if the manipulations are to be precise. Human perception of the luminance scale is neither simple nor linear. The essentials of the visual perception and its aspects to the practical image manipulation are also explained in detail.

First part (Sections 2 to 4) refers to the present state of knowledge in the pertinent fields.

Section 2 briefly summarizes the current knowledge about the physiology of the human visual system. The text is focused particularly to these aspects of vision, which are important for understanding of the image perception on a digital display device. The Section includes visual system anatomy, functionality of the spatial image perception and perception of the luminance scale and colors.

An introduction to the image sampling and resampling is in the Section 3.

The methods for reduction of the luminance range are described in the Section 4. The aim is to preserve most of the image's subjective appearance while reducing the luminance contrast. Two state-of-the-art methods are presented. Modern methods are non-linear, non-uniform and mostly influenced by human physiology.

Second part (Section 5) presents the author's contributions to the present state. A new algorithm for accelerated geometry correction is presented in Section 5.2. The method approximates the corrected image accurately and enables for very fast hardware realization. Section 5.3 presents acceleration of the bilateral filter. The filter is core of several state-of-the-art methods for dynamic range reduction (HDR tone-mapping). But the filter is also a performance bottleneck. The improvement of the resampling for display devices is introduced in Section 5.4. The method involves examination of the optimal distance from the display (Section 5.5). The optimal distance is compared to the user's visual acuity (ability to resolve details). The relation distance-to-acuity is then used for optimizing the

resampling filter (Section 5.6). The visual system is modeled with selected response of the neural cells. The whole image transfer is calculated and optimized via the resampling filter.

2 Present Knowledge About Human Vision Physiology

2.1 Basic Anatomy of the Eye

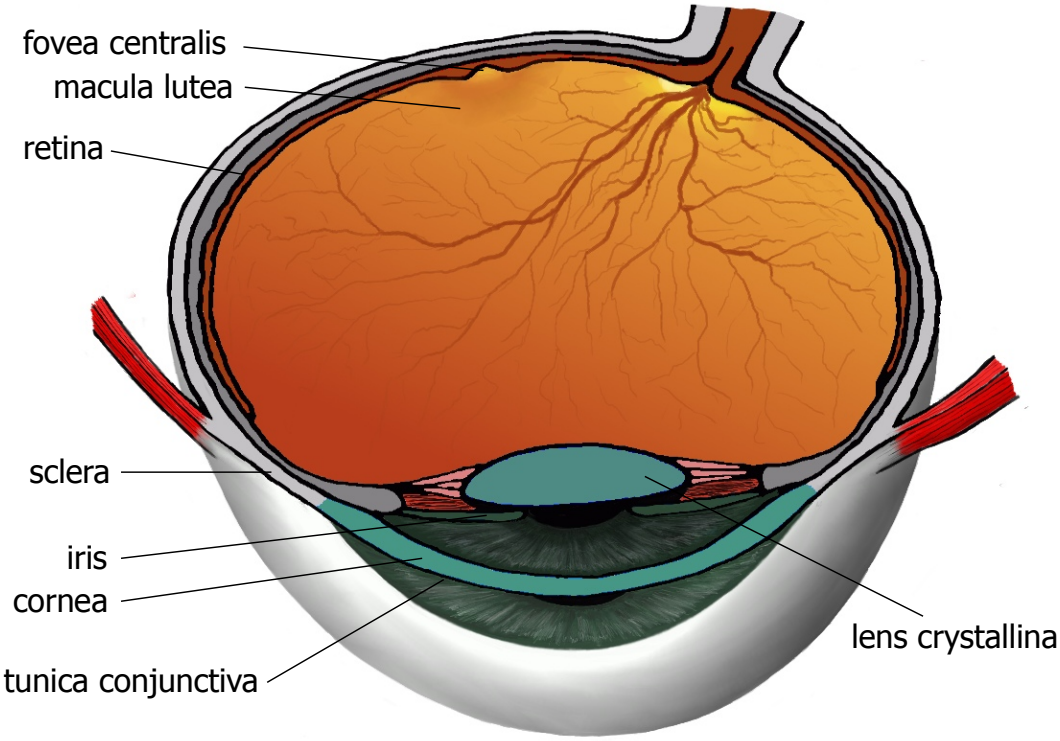


Figure 1: Transverse section of the eye, drawing by M. Seeman

As primates, we are together with some Cephalopoda equipped with the best kind of light sensitive organs in Animalia. Primate eye is equipped by complex optical part with several environments of different refractive indices, enables for accommodation, luminous flux control and three axis individual rotation.

Human eye is a spherical organ of rather exact size (25.4mm [3, p.3]). Six muscles turn the eye in three axes, horizontally, vertically and axially(around the optical axis) to expose it to a desired view. The optical part consists of conjunctiva, cornea, iris, lens and two chambers of liquid - aqueous chamber with aqueous humor and vitreous chamber with vitreous humor. The front convex curvature of the cornea performs two thirds of the total refractive power. The iris with pair of muscles changes the area of pupil. This way is controlled the amount of luminous flux. The lens, a soft, transparent tissue allows for accommodation. The tension to the lens edge changes the lens shape and thereby the refractive power. The proper accommodation helps to project the sharp image of the observed scene at the surface of retina, the light sensitive tissue.

2.2 Basic Anatomy of the Retina

The retina is highly vascularized layer of tissue, approximately 0.4mm thick. It contains photoreceptors and neural cells. The photoreceptors respond to the light stimuli and their signal is processed by several tens of specialized cell types [14]. Retina is full of neurons and therefore it is sometimes considered as a part of brain. Human retina is divided into central and peripheral. Central

retina (macula lutea) senses tiny details. Peripheral retina provides considerably lower spatial resolution, but is very sensitive to the motion. A typical observing of an object of interest consists of two phases. In first, the motion is detected by the peripheral retina. In the second, the eye is turned to the exact direction and stabilized (so called fixation). The object can be then examined by the central retina accurate vision.

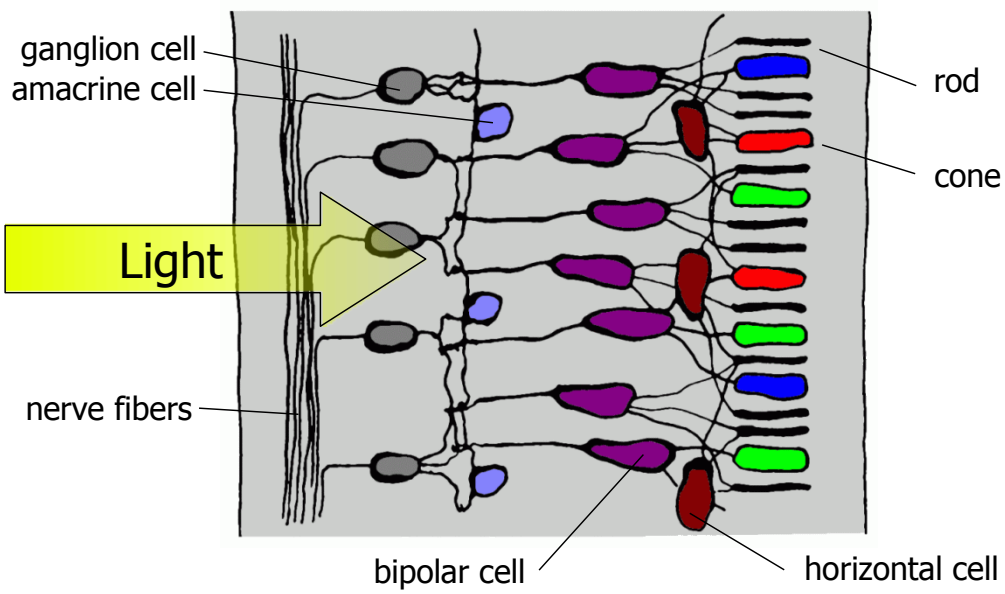


Figure 2: Basic scheme of retina neural connections

Photoreceptors

Two kinds of photoreceptors have been described, rods and cones. Rods are more sensitive and take part in nocturnal vision. Less sensitive cones work in daylight. Three types, L-, M- and S-cones in human retina are sensitive to different spectrum bands, which corresponds to the three basic colors - red, green and blue. L, M and S, means long, medium and short wavelength sensitivity.

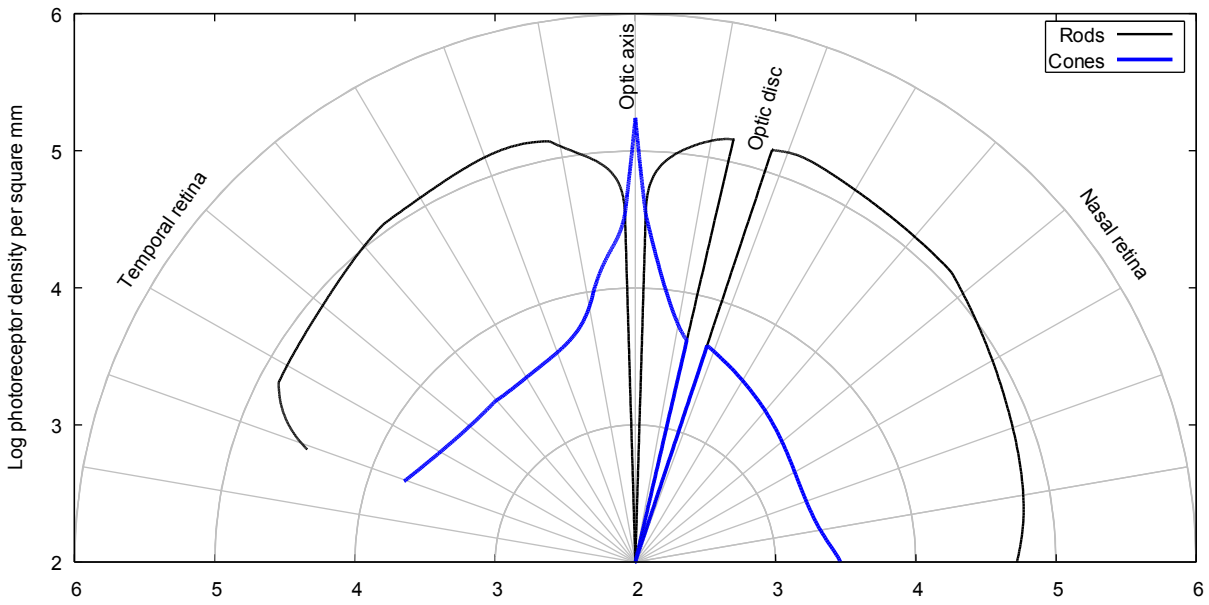


Figure 3: Photoreceptor density across transverse section [9]

Rods and cones represent two separate visual systems. The photoreceptors are separated by neurotransmitters, like dopamine and nitric oxide, released by specialized amacrine cells. When the retina is exposed to the high luminance level, neurotransmitters suppress the influence by the rod system [14]. This functionality controls photopic (daylight) and scotopic(nocturnal) mode of the retina. Hence the rod signal doesn't interfere visual sensation in high luminance levels. The cone signal is not suppressed in the scotopic mode, but they can not influence the nocturnal vision due to low sensitivity.

Fovea

The photoreceptors are not spread uniformly across the retina (Figure 3). As was mentioned above, human visual system is divided to central and peripheral part. Cones predominate in the central area. The very center of the retina is a specialized area called fovea (Figure 4). At the foveal pit there are no rods and the cone density is the highest of the whole retina. The fovea provides the finest vision and we use it to examine details. On the contrary, rods are dominating in the peripheral retina.

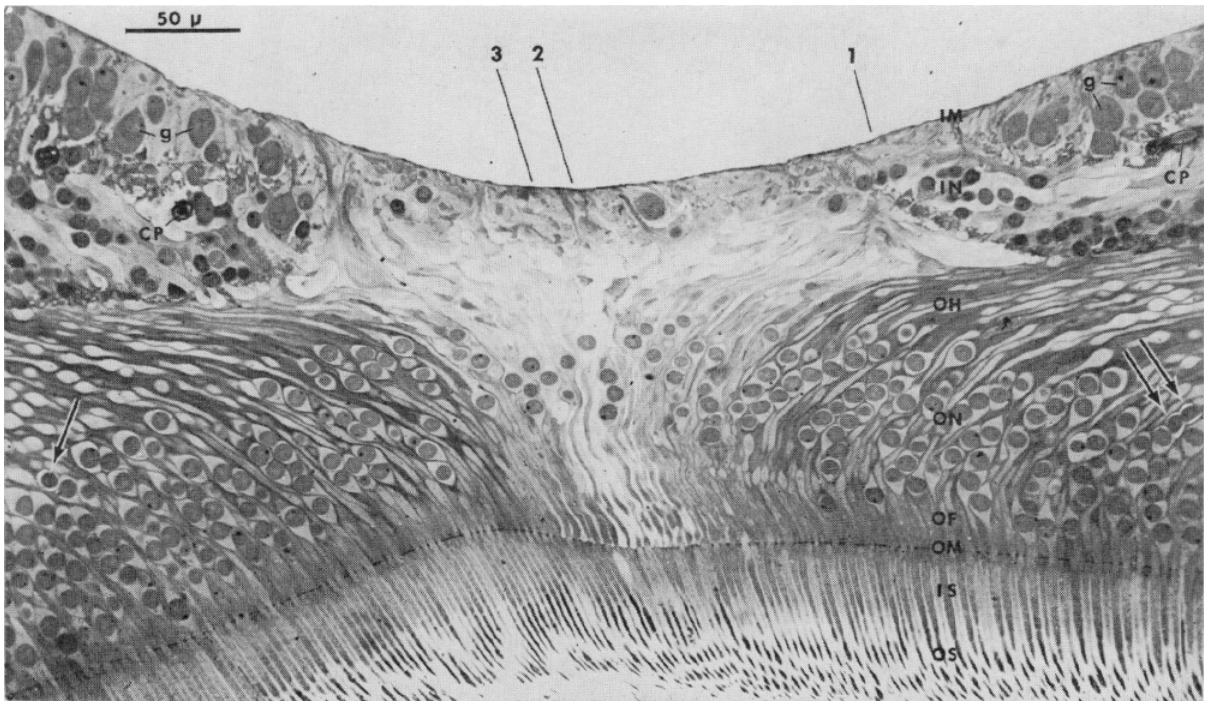


Figure 4: Human fovea section by Yamada [29], OF, IS, OS - *different part of the cone layer,* g - *ganglion cells*

Cell Signals

The retina neural cells form a complex structure. It can be divided into photoreceptor layer, outer plexiform layer, inner nuclear layer, inner plexiform layer and ganglion cell layer. More details can be found in [14 (ch. 2,12), 3]

The photoreceptors are in the very outer layer of the retina. The reason is a contact with the pigment epithelium, important for the photoreceptors metabolism. It follows, that the light, coming from the lens has to penetrate all layers of the retina before stimulating the photoreceptors. The photoreceptor reacts with a neural activity, which is processed by other specialized cells.

Recording the cells electric potentials enlightened the fundamentals of the retina network functionality [1, 2, 10, 11, 12, 13]. The cell signal form depend on the particular cell type. Some cells respond with *slow potentials*. On activation, they get gradually more hyperpolarized. This signal is recorded directly as voltage. The other cells respond with *action potentials*. Their activity is measured as a count of hyperpolarization spikes per second. The action potentials are important in transporting the signal to longer distance, where the voltage fades.

Retina Cells Receptive Fields

The information travels from the photoreceptors through several different types of neural cells before it reaches the visual cortex in the brain. By now, tens of cell types and connections are described [14] however one must keep in mind that the complete retina behavior is still not revealed by the physiologists.

Signal recording has revealed most of the cells function. The cell neural activity can be recorded either extracellularly or intracellularly. In extracellular recording, the electrode is placed to the cell proximity and records the activity of more cells. In the intracellular recording the electrode has to penetrate particular cell and recorded signal is not affected by other cells, e.g. [13].

A cell's spatial response is a function of spatial position in the cell's surroundings, where the retina is stimulated by light, and it refers how much response is evoked (the response unit depends on the particular cell type). The receptive field of different retina cell types vary a lot (see Figure 5).

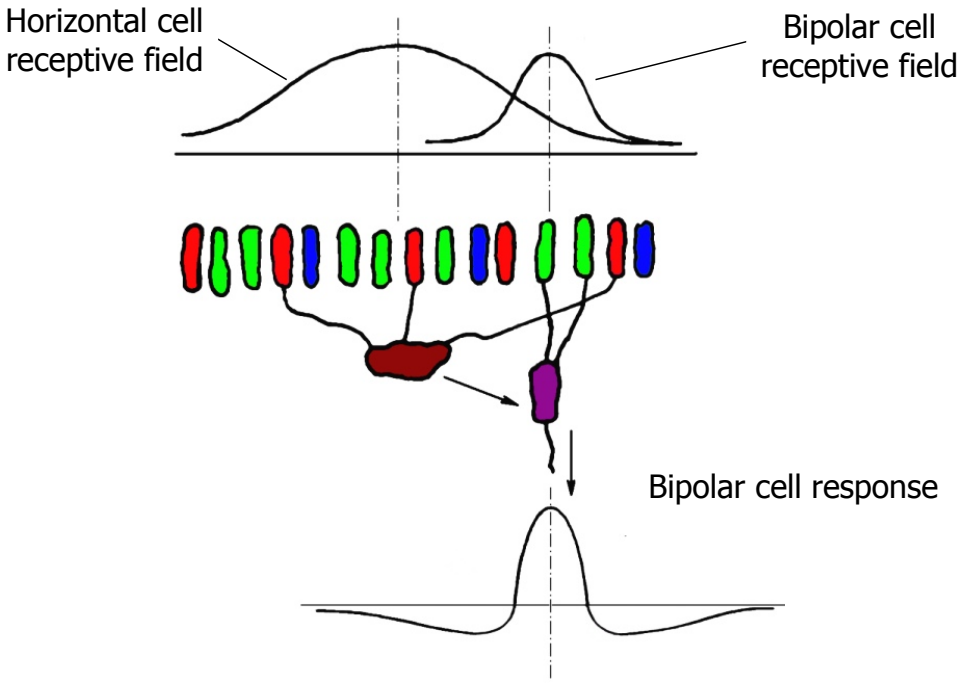


Figure 5: Bipolar and horizontal cells spatial responses

The essential principle of the human vision spatial characteristics is the signal fusion in horizontal and bipolar cells. Both types respond with the slow potentials. Horizontal cells collect the signal from many photoreceptors so their receptive field is rather wide. Moreover, their plasma membranes pass their potentials to the neighboring horizontal cells. Consequently their receptive field gets even wider. The signal travel across the membrane is strongly suppressed by neurotransmitters produced by amacrine cells when exposed to the light [14]. Therefore, when the receptive field is being measured, the average luminance has is an important parameter.

Bipolar cells on the contrary collect the signal from few photoreceptors, so their receptive field is very narrow. Some bipolar cells connected to a single cone only provide the narrowest receptive field possible. These cells take part in the finest central vision. The real response of the bipolar cells includes an inverse of a neighboring horizontal cell potential. The result is a difference between two Gaussian-like responses (see Figure 5). The plot is similar to the common high-pass filter response. The characteristic shape is similar in the spatial response of other cell types which process the signal after bipolar and horizontal cells. Even the ganglion cells, last neurons to process the signal before it travels to the visual cortex, have similar spatial responses, though they respond with action potentials. Recording of the ganglion cell signals is particularly important, because it refers to the response of the whole eye.

The receptive field of ganglion cells has been measured, mainly in cat, fish and monkey [1, 30], Figures 6, 7.

The ganglion cells in the retina are separated between parvocellularly projecting (P) and magnocellularly projecting (M) cells. While P cells are sensitive to fine, sharp but still details, M cells are sensitive to the motion.

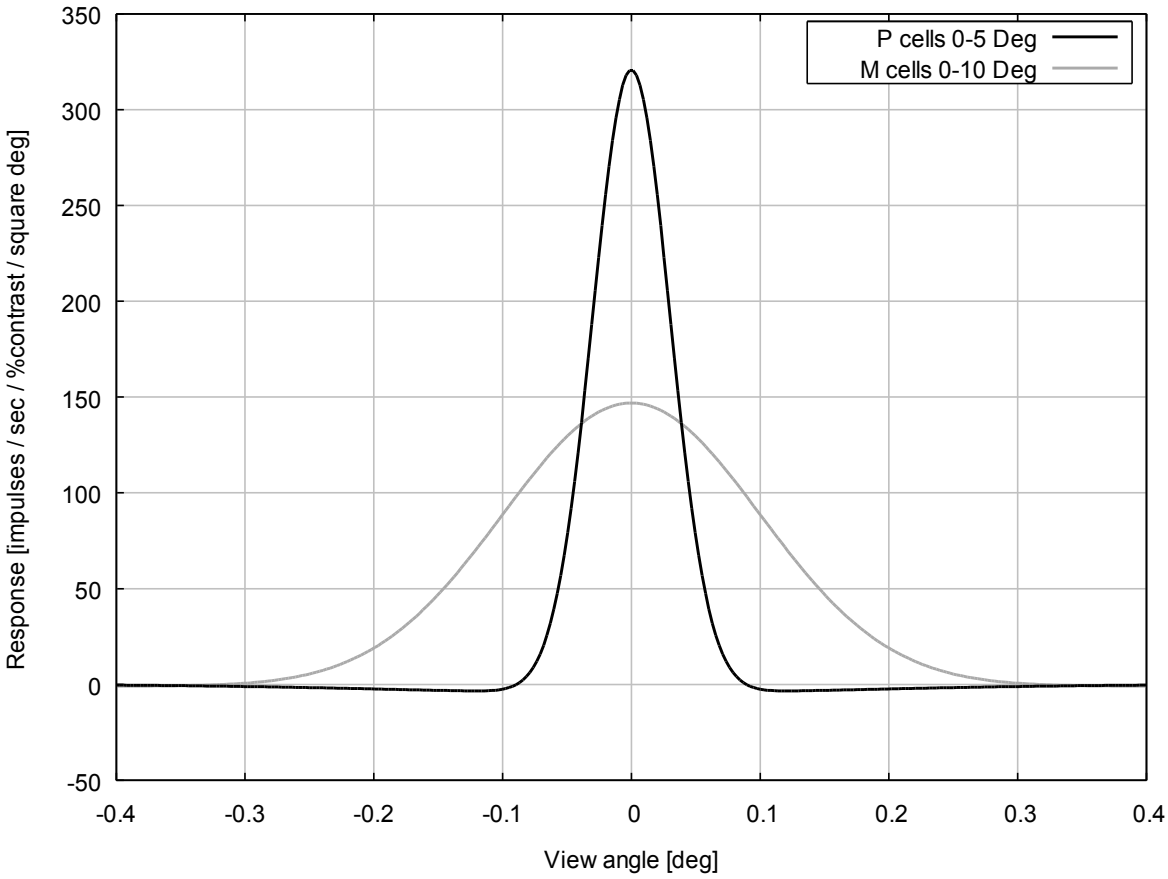


Figure 6: Section of ganglion cells spatial response in the central monkey retina, eccentricity: P cells 0-5 degrees, M cells 0-10 degrees [1]

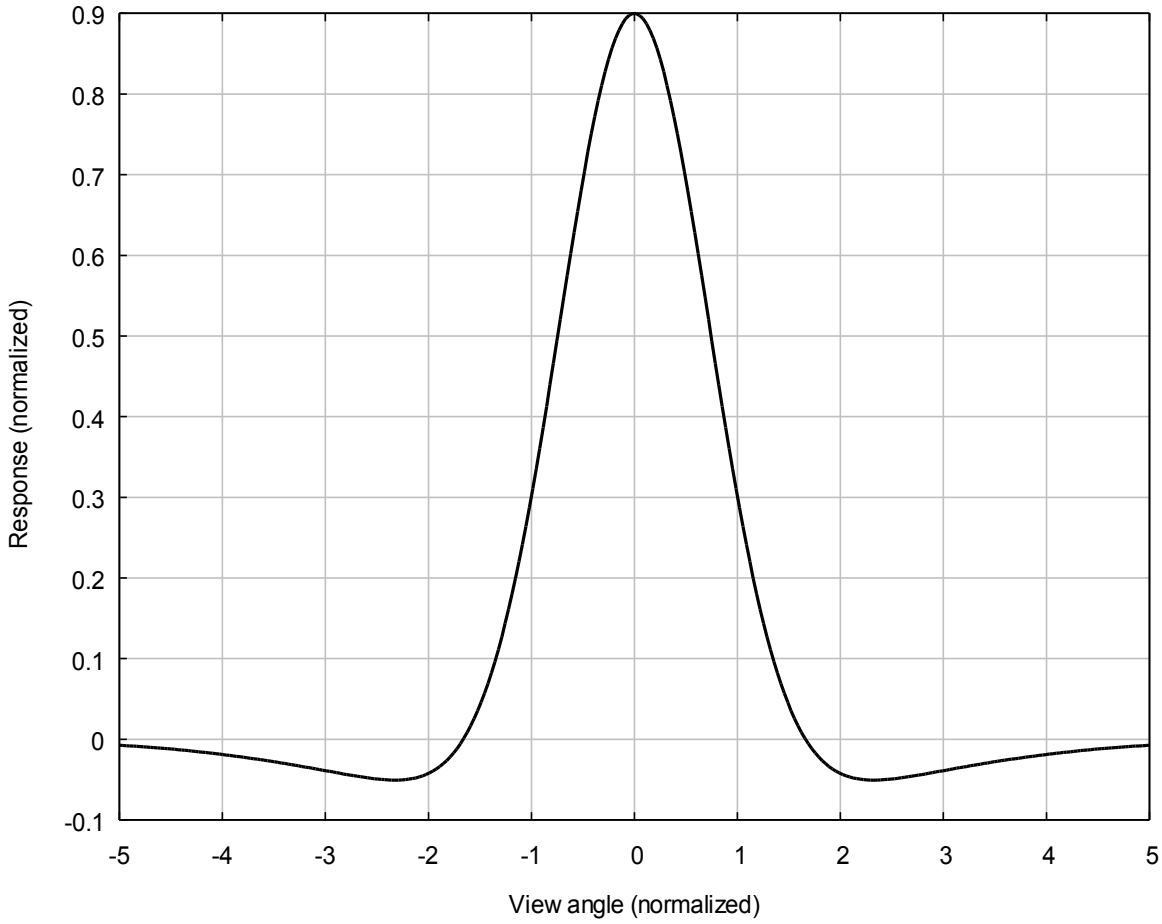


Figure 7: Spatial response of retinal ganglion cell in cat [30], 1D integral

Spatial Characteristics in Human

Mostly for ethic reasons, only the non-invasive measurement is available in human. Visual acuity [3 ch. 7] is the most common measure used by ophthalmologists. It describes few spatial vision features, however high spatial frequency cutoff due to optical aberrations and photoreceptor distribution is measured well. The subject is asked to resolve and describe different shapes observed from a specific distance. The smallest shape or the shape with the finest details just resolved gives the result value. Visual acuity changes with luminance, therefore distinct illumination level in the test-room may be demanded. Alternatively the illumination should be measured as a part of the result. Several kinds of acuity were described.

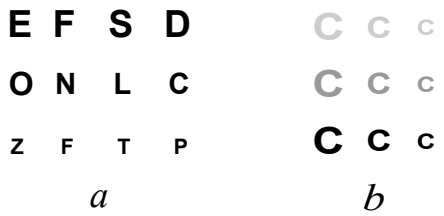


Figure 8: Visual acuity optotypes; a) Snellen acuity chart, b) Bailey & Lovie optotypes

Contrast Sensitivity Function

Contrast Sensitivity Function (CSF), [17, 3 ch. 7], Figure 9 is another commonly used non-invasive measure. It refers to the HVS response to different spatial frequency in the observed scene (the term spectral sensitivity may be confused with the color spectrum).

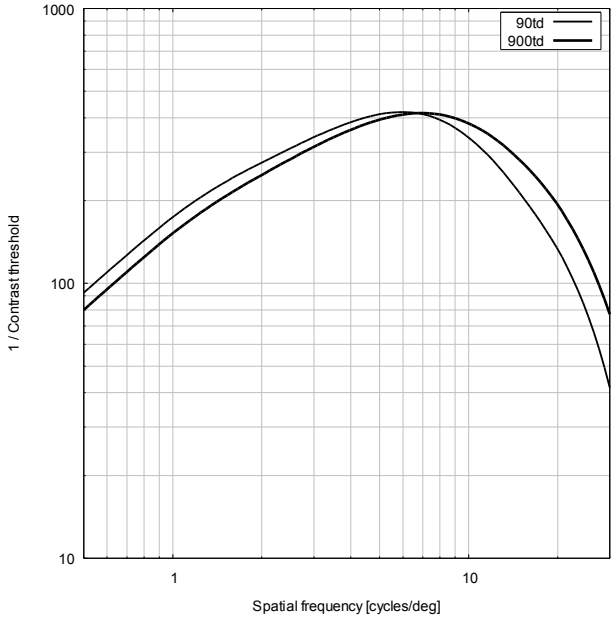


Figure 9: Contrast Sensitivity Function [17]

It involves exposing a graphical patterns with different spatial frequencies to the subject. It is difficult to measure subject's precise sensation, so the contrast of the pattern is changed gradually. The particular frequency response is determined from the contrast needed to just resolve the pattern. The human CSF is a typical band-pass filter response [16], the luminance level of common displays corresponds to the retina illumination approximately between two selected curves (50 and 500 trolands[1]).

The high frequency inhibition is mainly due to the optical aberrations in the cornea and the lens. Even though the sight can be perfected by an individual correction to the supernatural acuity [21], a resolvable detail is still limited by the size of the photoreceptor.

On the contrary, the low frequency inhibition couldn't be explained by the eye's optical structure. The smooth patterns are suppressed via the neural network in the retina.

2.3 Luminance Scale

The perception of the luminance scale uneasy to describe. The signals from several kinds of neural cells do affect each other. Some signals are modulated by neurotransmitters. Many of this mechanisms help HVS to perceive the vast range of the luminance. Fast and global adaptation is provided by the iris. But unlike in the common belief, iris does not perform the major luminance adaptation. Most powerful luminance adaptation is the photopic/scotopic mode control. The full adaptation lasts several minutes, exact time characteristics were modeled in [5]. Other form of light adaptation is the signal processing performed by the retina neural network. It seems that the retina signals refer more to the contrast between various areas than to the absolute luminance.

1 Troland (Td) is a retina illuminance unit. One troland is the illuminance of the retina by an external luminance 1cd·m^{-2} through a pupil with an area 1mm^2. Note that troland differs for various genus, because the definition depends on the eye diameter, or more precisely, on the distance pupil-to-retina.

It is commonly known, that the dependency between observed luminance and the psychical sensation is close to logarithmic. Even if we look over the fact, that psychical sensation is difficult to measure and compare the observed scene with the ganglion cell signal, the findings are more complex.

Purpura & Kaplan [2] have measured the dependency of the ganglion cell activity on the mean scene luminance and local contrast in monkey.

The stimuli were sine gratings of different mean luminance and contrast. The sine spatial wavelengths were chosen to correspond to each ganglion cell's maximal response. As ganglion cells respond with action potentials, the response is measured as the hyperpolarizarion frequency in the recorded signal.

Figure 10 shows that the response dependency on the local contrast of the sine gratings is almost exactly linear. We were also looking for the response dependency on the global illumination. In Figure 11 the response is plotted as a function of absolute illumination contrast (L_{max}-L_{min}). The sections of constant contrast percentage are close to logarithmic curves (close to lines in the chart, as the abscissa scale is logarithmic) Note that in our charts the illumination values are recalculated to human trolands according to [2].

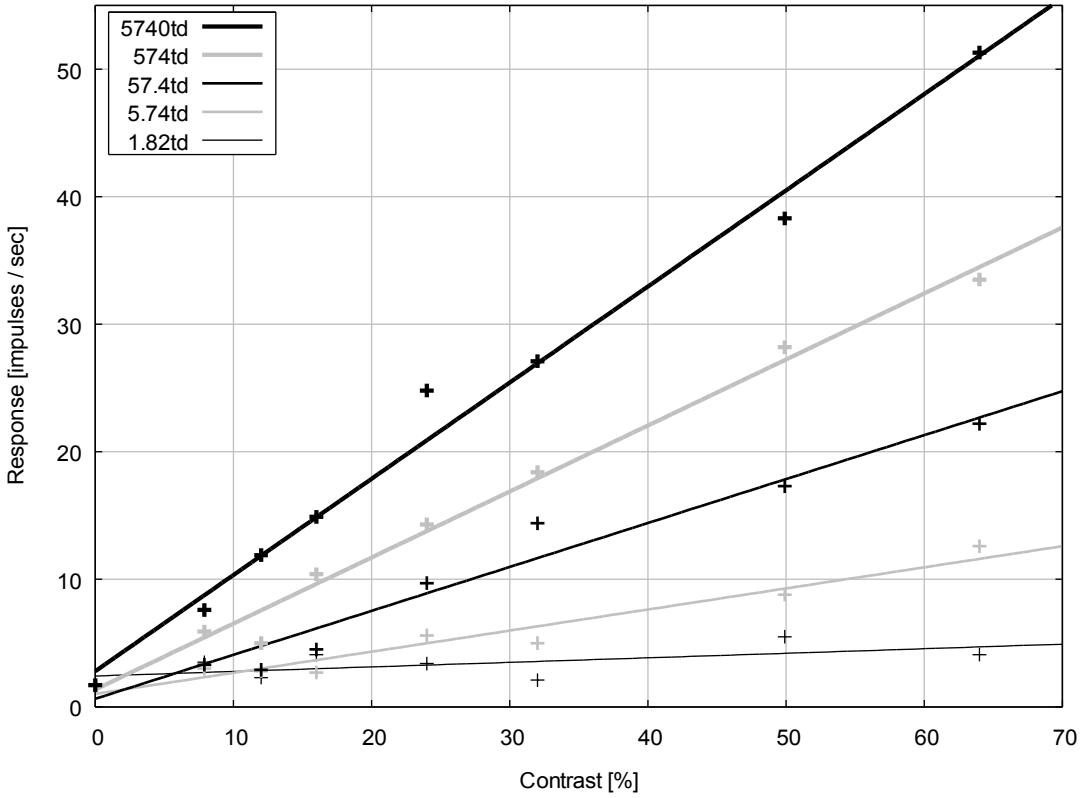


Figure 10: Ganglion cell response dependency on local contrast and global illumination in monkey[2]

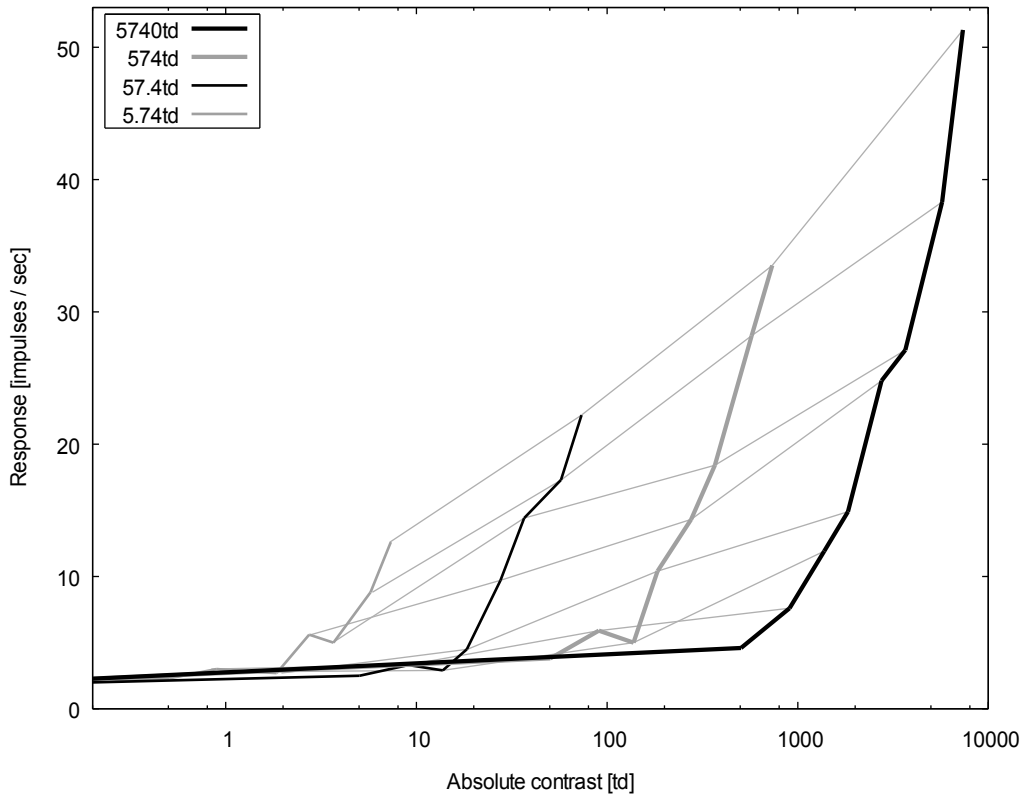


Figure 11: The same dependency [2] plotted as a function of absolute contrast in trolands

Color Perception

There are two kinds of color duality in the eye's neural signal. Red-Green and Blue-Yellow. The color-opponent signals emerge firstly in the bipolar cells which combine positive and negative s-potentials of different cone types [22]. The color opponency was discussed and refused for a long time. Present knowledge states, that most of the ganglion cells are color opponent [35]. These pathways show some form of time persistence and are sources of the famous afterimages. When the eye observes a still color image for a while and then stares at the neutral gray area, the color opponent image is observed.

The important fact is that the brightness information is passed separately and the color channel acuity is approximately one third of the brightness channel acuity [18]. Unfortunately it does not mean that the hue perception is not dependent on the luminance. The dependency is described as the Bezold-Brücke phenomenon [15].

The same ratio of the base colors is interpreted slightly differently by the HVS when luminance level changes. The subjectively perceived hue seems to turn closer to green and red with the raising luminance level. Four different hues, so called constant hues, appear not to change with the changing luminance. Yet the hues are constant over a small luminance range and with very different luminance level, either the “non changing hues” are shifted (see Figure 12).

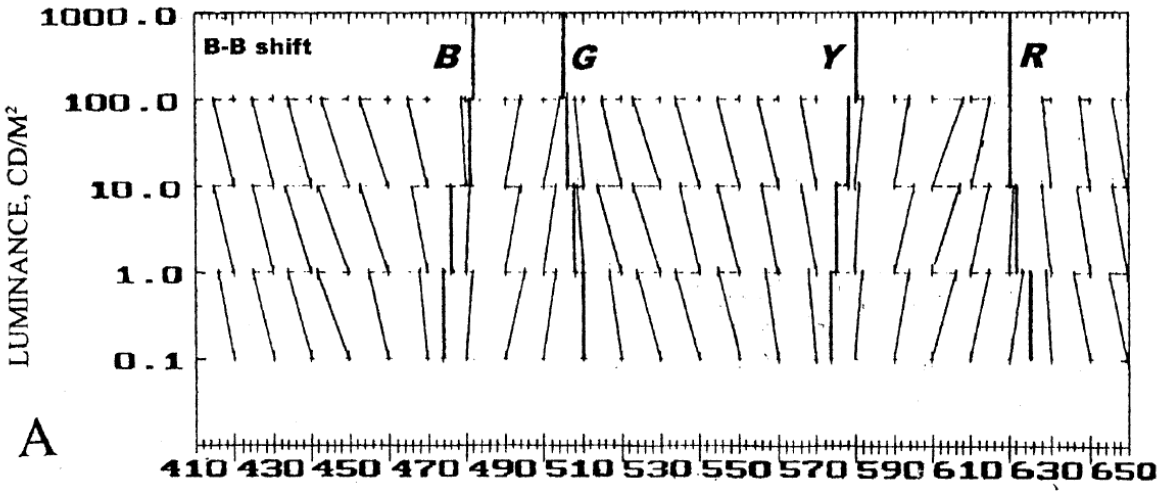


Figure 12: Bezold-Brücke Hue-Shift [15] in the extended wavelength scale [23] reprinted with kind permission of Dr. Pridmore

3 Present State in Sampling

The ideal image subsampling in signal processing is clear: According to the Nyquist theorem the sampled signal can not contain any harmonic wave of higher frequency than one half of the sampling frequency. Therefore the function (represented by the original samples) has to be filtered with a low-pass filter, which removes any harmonic wave unsuitable for the lower sampling rate (see Figure 13)

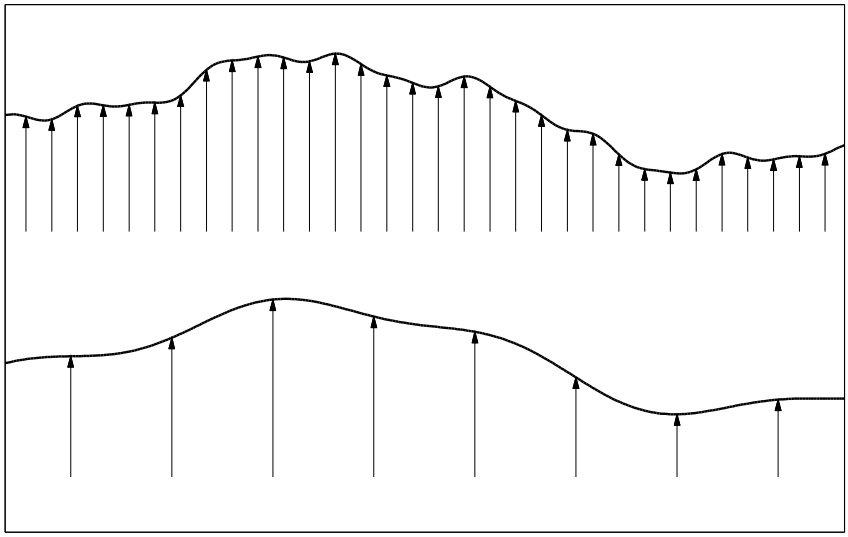


Figure 13: Subsampling illustration

The ideal low pass filtering is a convolution with a properly scaled sinc function. For the computation complexity the sinc function is mostly windowed in practice. One example is Lánczos filter (see Figure 14), which's accuracy is very close to optimal and computing time is significantly shorter.

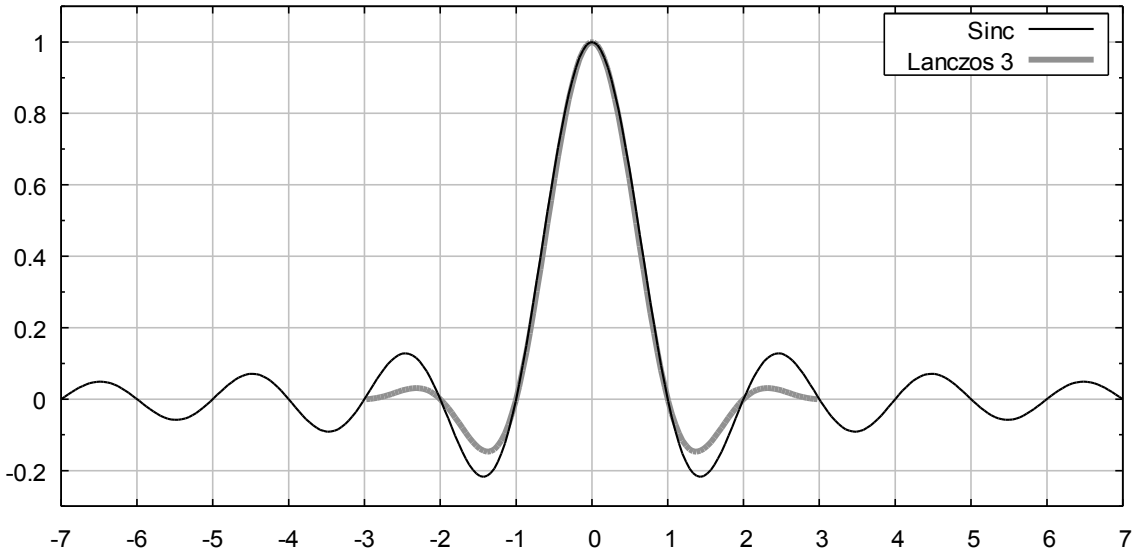


Figure 14: Sinc and Lánczos filter

If the situation was considered only as a transport of the image to the display, there would be no space for improvement. But HVS is a part of the whole perception process. And HVS changes both the

spatial sensation and the luminance scale. The whole system can not be optimized by optimizing only the part of the system, which resamples the image.

As for the color information, many of the image processing algorithms do assume that the R, G and B luminance channels carry an independent information and therefore they can be resampled separately. Even this is true only when HVS is not involved. Hue perception is affected by luminance level. Details are explained in the following Section.

4 Methods for Luminance Scale Displaying

Tone-mapping of the high dynamic range images became very topical in recent years. Tens of new algorithms were presented. The HDR displays are still not massively available, and using of HDR tone-mapping is still actual.

In general, the tone-mapping operators (TMO) can be divided into two groups. One group is of physiologically influenced operators. The others aim is only to produce a visually attractive image. The relevance to the original scene or to the subjective perception of the scene is not important. For this work, the first group is more important and so the latter will not be discussed here.

Two important approaches have been selected for this Section

Dynamic Range Reduction Inspired by Photoreceptor Physiology

The main point of this article [33] is modeling the nonlinearity of a photoreceptor. Unlike in our survey (Section 2.3) authors use the nonlinear luminance characteristics of the cones only.

The cone potential dependency on the intensity I is modeled as

$$V = \frac{I}{I + \sigma(I_a)} V_{max} \tag{4.1}$$

...where I_a is the adaptation intensity and adaptation function σ is given by

$$\sigma(I_a) = f(I_a)^m \tag{4.2}$$

f and m are parameters of the operator.

Adaptation intensity is calculated as a linear combination of the pixel intensity and mean intensity in the image.

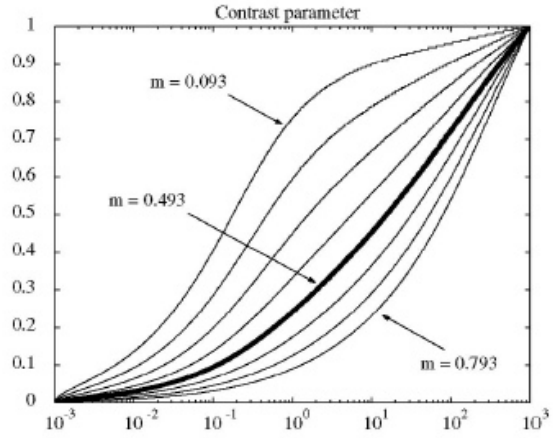


Figure 15: [33]: Mapping of input luminances (horizontal) to display luminances (vertical) for different values of m.

One advantage of this method is, that the three cone types can be modeled separately by adjusting of the parameters.

A Local Model of Eye Adaptation for High Dynamic Range Images

This method [5] tries to implement most of the known luminance dependent features in the human retina. The aim was obviously not performance but precision. The modeled features include:

- Precise time adaptation
- Both photopic and scotopic modes, modeling
- Light adaptation is precisely computed by bilateral filtering of the image

The response is computed similarly to the previous method.

$$R = \frac{Y^n}{Y^n + \sigma^n} \tag{4.3}$$

Also the adaptation function is similar

$$\sigma(I_a) = \beta I_a^{\alpha} \tag{4.4}$$

… where α and β are constant, different for rods and cones and n is the sensitivity constant. The response is computed separately for photopic and scotopic mode, Y refers to the particular luminance.

When the luminance some area is in a mesopic range (between photopic and scotopic), the response is calculated from both types of photoreceptors by so called mesopic factor.

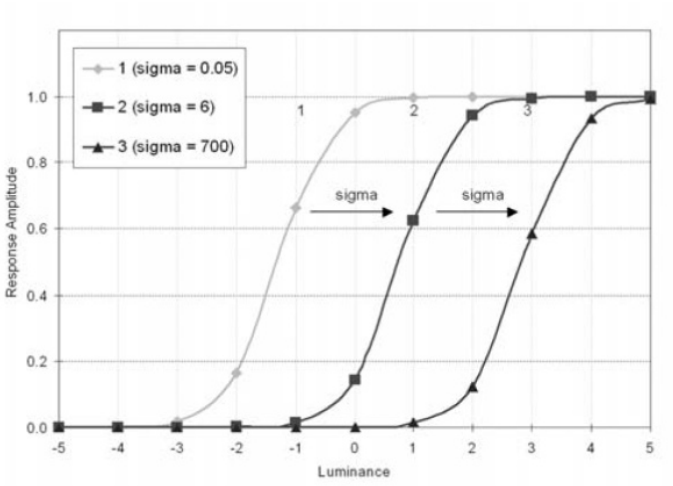


Figure 16: [5]: Adaptation function

5 Perception Optimizing on a Display

This part is about practical contributions to the image reconstruction on display. After a brief description of the whole setup, the particular improvements will be explained in detail in the following Sections.

5.1 Overview

Let us assume that display contains a matrix of horizontally and vertically aligned rectangular units (pixels) capable of emitting a light of different luminance and color.

Then let us state that an image represents information about the luminance distribution across a scene.

Rendering an image on a display comprises of following problems:

1. Inaccuracy in the scanning. This covers geometry distortion, brightness and hue bias. These aberrations ought to be compensated.
2. The image is almost always adjusted. Let us put aside the fact that users may want to modify the image. There are unavoidable changes because of display imperfections. The dynamic range of the common devices is not sufficient for scenes with high contrast and 8-bit quantization is too rough for scenes with low contrast. In most cases, the dynamic range of the image is scaled to fit the display range. Change of the luminance may cause different hue perception (see Section 2.3, part Color Perception). In serious image reproduction, this should be also corrected.
3. Resampling to the device native resolution. This part should also include the shape of the pixels, as will be shown further below.

The physiology paper [25] proposes a framework for image reproducing and manipulation. In general case covered by this work it is extended by scanning correction and resampling (Figure 17).

Scanning correction includes geometry correction and range scale correction. Neither is difficult to perform with sufficient source data. However the geometry correction may be too slow in some applications. For this purpose, an acceleration is presented. The method was designed specifically for fast processing of images distorted by optics, where the displacement distance is not long. The method is described in Section 5.2 Image Resampling for Geometry Correction.

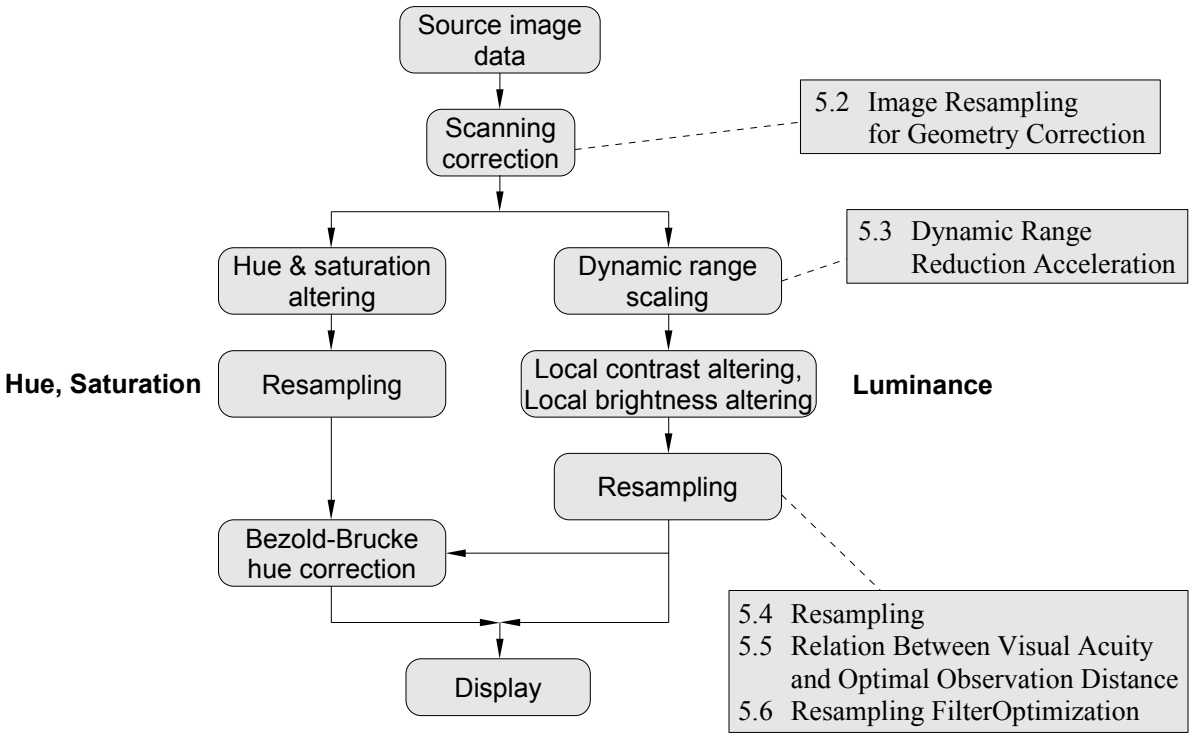


Figure 17: General framework for image reproducing and manipulation

Methods for dynamic range reduction may be performed in several ways. Very simply by linear scaling of the luminance (this is equal to the change of the illumination) or by scaling of logarithmic range. The scaling of gamma shaped luminance is incorrect, it follows neither physical nor physiological behavior. More complex alternative is using of some more complex HDR tonemapping operator, e.g. dividing by bilateral filtered image [5, 6, 7] (more of luminance scale processing was described in Section 4). The weakness of these state-of-the-art methods is the bilateral filter computation complexity. A close and accurate approximation for bilateral filter designed for especially for HDR tone-mapping is presented in Section 5.3

In the scheme in Figure 17, the adjusting is split between brightness and color. The reason is partly noticeable from the physiology (see Section 2.3, part Color Perception). In the HVS the pathways are separated. Also the resampling is separated to the brightness and color branch. This needs closer explanation. In theory, the resampling should be independent on all range operations and therefore it might be placed anywhere in the above scheme. Furthermore, either the geometry correction is clearly a resampling. But now the matter is the final resampling to the display device, which can not be error-less. This is due to the size of the pixels. The pixel shape is recognizable by human eye, so an error is brought into the image. It means that the signal is changed and the resampling may be treated as an altering operation. In this case the brightness resampling is the last operation before the display device, but the color resampling is followed by hue correction. It is correct, because errors in the brightness and hue resampling are reflected in the hue correction. The Sections 5.4 to 5.6 describe the resampling optimization.

The luminance adjusting is divided into local contrast adjusting and local brightness adjusting. This corresponds to Section 2.3 Luminance Scale. Luminance range perception is different for contrast and mean illumination.

Local contrast altering, e.g. edge enhancing or smoothing filter applies for small areas, mostly single pixels. The sensation is linearly dependent on luminance difference between the pixel and its surroundings. Besides, if a filter of this kind is applied in any but linear scale of luminance, brightness

would be changed (see Figure 18). The reason is the physical nature of light. Luminance in any optical system is integrated in linear scale.

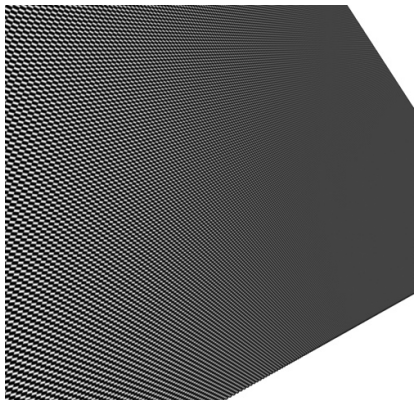


Figure 18: A black and white texture lo-passed in a nonlinear scale. The far area appears darker than the near, visible pattern.

Local brightness altering applies when the changed area is larger. The sensation depends on the absolute luminance value and the dependency is logarithmic. A typical example is adding brightness to some part of an image. The area border should be soft, e.g. by using some window function, otherwise the local contrast on the area border would be affected too. An example of correct and wrong adding of brightness is shown in Figure 19. Adding a luminance in linear scale results in subjective contrast reduction. Again, operating with gamma shaped luminance scale is incorrect, though it is frequently used and the error is less significant.

a *b* *c*

Figure 19: Brightness altering of an image part. a - Original. b - Multiplying luminance. c - Adding luminance. The scale is calculated for observing via display / prints with gamma correction 2.2

If the Bezold-Brücke hue correction takes place, it should change the result hue to the hue, which would be observed under the original luminance. It means that both source and result luminance range must be known. But the absolute luminance value is not as important as the luminance range. The reason is that the perception of color adapts with time, so color differences across the image are subjectively much more apparent than the hue shift of the whole image.

The correction should be considered for each particular application. In scanned images with luminance range similar to the range of the display, the correction may be insignificant. The transform should be applied when the source image luminance range is much greater than the display luminance range. In this case the scene range has to be reduced, which causes subjective perception of a different hue.

With synthesized images it is important whether the source data represent physical color or subjective hue. The correction for subjective hue is straightforward, the absolute hue shift for each

pixel luminance has to be used. The correction for physical source data is same as with the scanned data, only the hue shift caused by the luminance difference is counted (see Figure 20).

The last part which has not been mentioned is the resampling in the hue & saturation branch. The problem is shortly described within Section 5.4 Resampling, but it is not trivial and will be not solved in this text. Instead, the optimization of the color resampling is planned as a future work.

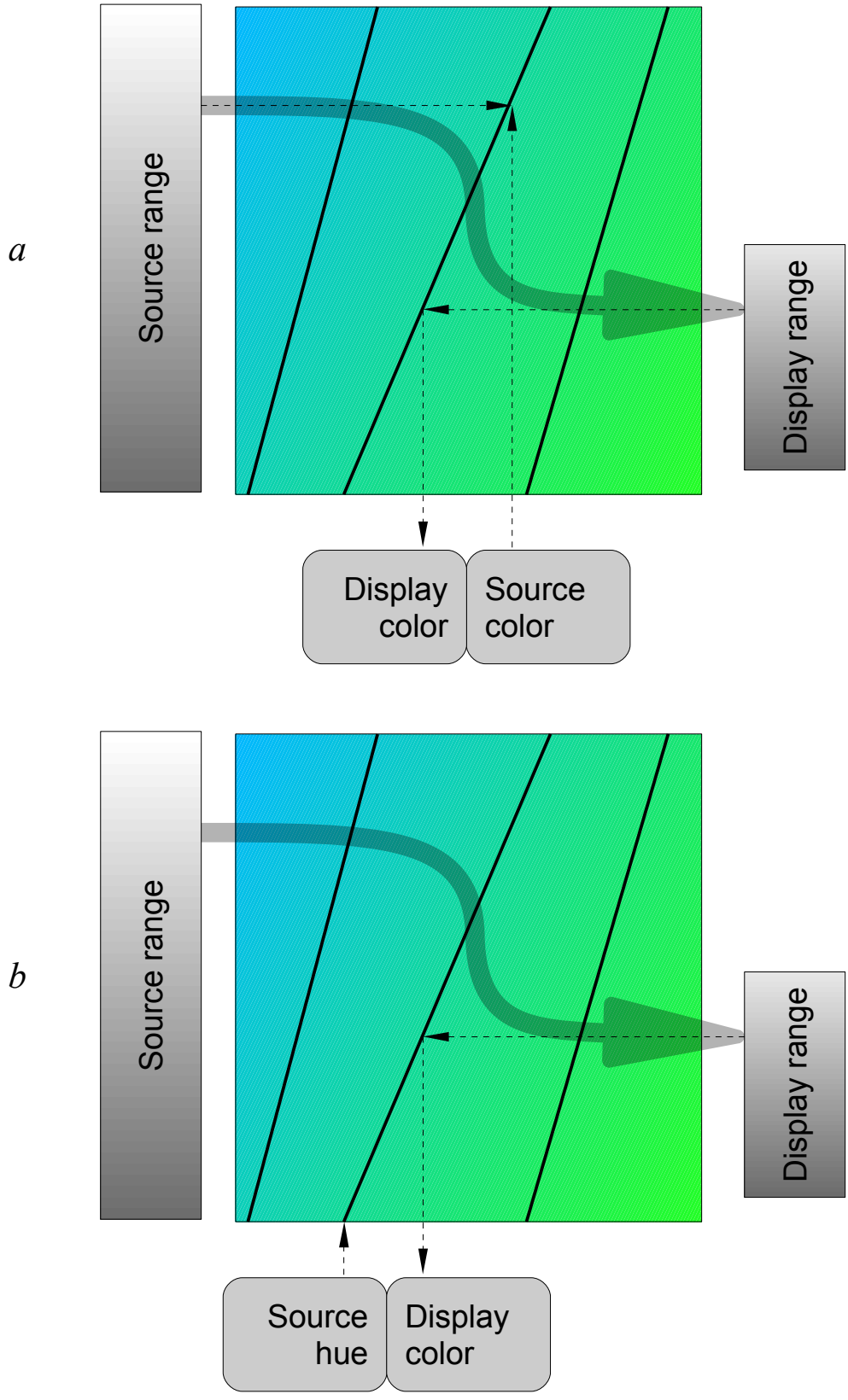


Figure 20: Using Bezold-Brücke correction. a – scanned data or synthetic image representing physical color, b – synthetic data representing subjective hue.

5.2 Image Resampling for Geometry Correction

The geometrical distortion may be unacceptable in some applications. Therefore it is desirable to acquire geometrically correct image. The presented algorithm helps in correcting such images. The algorithm provides high performance at the price of certain limits. The displacement and rotation should stay in some constraints.

The geometrical correction itself - calculation of new sample positions corresponding to the new image samples in the original image - is, in the presented case, relatively straightforward. It can be performed through approximation of the distortion through rectangular mesh placed over the image where displacement in the corners of the mesh is known and within the nodes it is calculated through bilinear interpolation of coordinates.

The main problem of geometrical image correction is getting the new samples values so that the signal properties of the image remain as much preserved as possible. The often used methods, such as the nearest neighbor method, which heavily damages the image signal properties, and bilinear or bicubic interpolation which can be better, but by far is not ideal, are traditionally used for this purpose. The main reason is that while the algorithms preserve good signal properties, namely frequency spectrum, are known, they are often considered prohibitively computationally expensive.

The presented approach uses a resampling filter that is far better from the point of view of signal properties than the bilinear or bicubic interpolation while still preserves relatively low computational requirements and allows for real-time implementation. The filter used in our approach is finite impulse response filter (FIR). The main advantage and novelty of the proposed approach is the combination of simple yet efficient filter structure combined with efficient implementation. The potential drawback of the approach is that it can be used only in cases where the distortions do not involve significant angular or scale changes. The method allows for several pixels displacement with sub-pixel resolution and angles limited to less than $\pi/20$ [36]. This, however, suits exactly the purpose of the above-mentioned applications.

The resampling algorithm can be split into two relatively independent parts: calculation of sample (pixel) values for the known sample location in the original and calculation of the location (displacement) of the pixel locations.

First of all, let us summarize the requirements and properties of the resampling in the presented case, as indicated above:

- The displacement of the pixel location is small.
- The angular distortion is negligible.
- No scaling is involved in the correction.

The proposed selection of algorithms is based exactly on these properties. Although the choice can reduce generality, it leads into very efficient implementation as shown below. In this text, it is also assumed (without further loss of generality) that the image is a single-channel image - every pixel has a single scalar value belonging to some set of values (e.g. integer number).

Sample Value and Separability

The general task of resampling is merely to get precise sample values for every sample (pixel) of the output image based on the knowledge of the pixel displacement and, of course, ensure that the limit frequency of the sampling theorem is respected. From the above properties of the algorithm, it can be concluded that the sampling theorem cannot be violated through resampling (the sampling frequency remains the same in the input and output images) so the precise value of the sample is the only concern. The task of reconstruction of a sample value can be described through the following equation:

$$R = f(O) \tag{5.1}$$

where R is the resampled image, O is the original image, and f is the resampling function. In many cases, the above function can be made separable, so that its calculation can be split into two independent parts. As the separation allows in efficient implementation, the approach was also used in the presented case.

$$R = f''(f'(O)) \tag{5.2}$$

where R is the resampled image, O is the original image and f' and f'' are the separated resampling functions. The functions f' and f'' can be implemented using several methods. In the presented approach, the finite impulse response (FIR) filter implementation was chosen as the general principle for the value reconstruction functions. The preservation of the signal properties is highly dependent on the size of the neighborhood of the pixel (in the theory, the signal can be fully reconstructed only from the complete set of samples), for practical purposes, neighborhood of several pixels is enough. The resampled signal can be evaluated using the rewritten equation.

$$R = FIR''(FIR'(O)) \tag{5.3}$$

Where FIR' and FIR'' are the separated resampling functions. The selected solution that uses FIR filters has an additional advantage in the fact that the FIR coefficients can be easily exchanged and therefore the features of the sample reconstruction can be fine-tuned based on the application, if necessary at all. The filters considered further in this paper are Lánczos filters (see Section 3) Note, that in the FIR functions are dependent on the location (i.e. filter coefficients in this notation would have to be calculated separately for each output sample).

Geometry Correction

To enable proper reconstruction of the output sample, it is necessary not only to have proper sample reconstruction function (as described above), but also the accurate location of the desired sample in the original image. The location of the sample is possible to evaluate through an absolute location, but for the purpose of the presented approach, it is better to evaluate the displacement (difference between the location of the sample in the original image and its place in the output image) for each pixel, see Figure 21 for illustration.

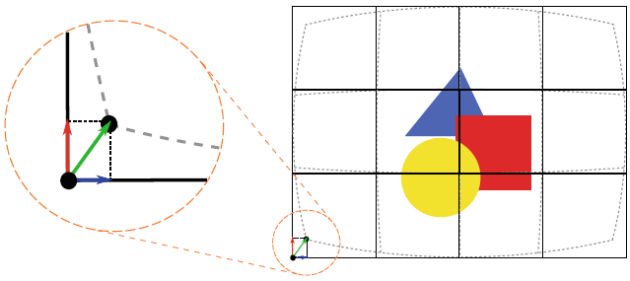


Figure 21: Fig. 2 Displacement between corresponding pixels in distorted original image and output image marked with green arrow. The displacement can be decomposed into vertical displacement (red arrow) and horizontal displacement (blue arrow)

$$(x_o, y_o) = (x_r, y_r) + d(x_r, y_r) \tag{5.4}$$

where (x_r, y_r) is the pixel location in the resampled image, (x_o, y_o) the location in the original image, and d is the displacement function. The displacement can be also made separable, e.g. as shown in the equation below.

$$(x_o, y_o) = (x_r + d_x(x_r, y_r), y_r + d_y(x_r, y_r)) \tag{5.5}$$

where d_x and d_y are the separated displacement functions. To enable two completely separated steps in location evaluation, it is also possible to rewrite the above equation.

$$(x_o, y_o) = (x_r, y_r) + (0, d_y(x_r, y_r)) + (d_x(x_r, y_r), 0) \tag{5.6}$$

In the presented approach, the correction function is done through bilinear interpolation within pre-defined rectangular mesh (see Figure 22). Location in the original image can be expressed through the following equation.

$$\begin{aligned}(x_{o,S}, y_{o,S}) &= (x_r, y_r) + (0, d_{y,S}(x_r, y_r)) + (d_{x,S}(x_r, y_r), 0) \\ &= (x_r, y_r) + (0, y_r + p' x_r + p'' y_r + p''') + (y_r + q' x_r + q'' y_r + q''', 0)\end{aligned} \tag{5.7}$$

The above described algorithm can be implemented using a differential scheme, where for each of the rectangular nodes, a set of four pre-calculated coefficients can be used:

- D0 top left pixel displacement.
- DC0 difference of displacements between adjacent pixels in first row of the square.
- DR difference of displacements between first pixels of adjacent rows.
- DDC change in difference of displacements between pixels of adjacent rows, that means $DC_{n+1} - Dc_n$.

For more detailed description, see Figure 22. The displacement calculation can be subdivided into independent calculation of vertical and horizontal displacements. Pseudocode of the interpolation algorithm follows.

```
DoR = D0
DC = DC0
for each row in square
{
  D = DoR
  for each pixel in row
  {
    Output FIR[fp(D)](O,ip(D))
    D += DC
  }
  DoR += DR
  DC += DDC
}
```

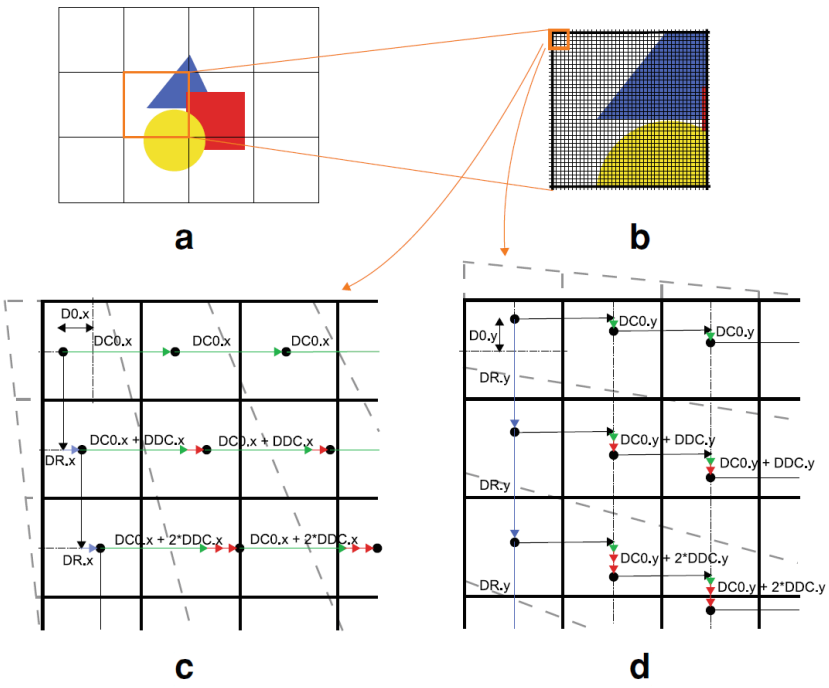


Figure 22: Displacement interpolation in squares. Pixels of original distorted image are plotted with gray dashed line, pixels of output image are plotted with black solid line. Meaning of precalculated coefficients is marked with coloured vectors

Complete algorithm

To efficiently implement the separated FIR filters and separated location evaluation, the FIR coefficients need to be pre-calculated and stored. For this purpose, the subpixel position of the location, on which the FIR coefficients are, in fact, dependent, can be quantized and used to index a table of pre-calculated FIR filters. So single samples filtered through a FIR filter, can be evaluated e.g. in the following way.

$$FIR(I,x,y)=FIR_{fp(x)}(I,ip(x),y) \tag{5.8}$$

where $fp(x)$ is a quantized fractional part of x, $ip(x)$ is an integer part of x, FIR_q is a filter from a filter bank precalculated for sub-pixel position q. (y is assumed to be integer number in this case). Therefore, the complete evaluation of a single pixel of the output image is as shown below.

$$r_{x,y}=FIR''_{fp(y+d_y(x,y))}(FIR'_{fp(x+d_x(x,y))}(I,x,y+d_y(x,y)),x+d_x(x,y),y) \tag{5.9}$$

where x and y are the integer coordinate representations, d_x and d_y are the displacements (functions of x and y). Finally, let us summarize the features of the algorithm. The algorithm is based on the FIR filtering which is suitable to preserve signal properties as well as possible with the option to fine tune the features according to the application. The filter coefficients should be pre-calculated into a filter bank dependent on sub-pixel position of the samples. The approach is also based on the assumption that the suitable filter is separable. The geometry correction is performed using simple bilinear interpolation.

Results

This approach enables for implementation using a pipeline with low consumption of resources in a programmable hardware (FPGA). Although the implementation proposed in the presented approach is simple, it preserves the image, as well as the more complex implementations of the filters given the constraints of the approach are respected. In [34, 37], the device design is also described. The described device clock frequency [34] is up to 105 MHz. While the resampling unit produces one output pixel per 2 clock cycles, the output resampling data rate is up to 52.5 Mpixels per second.

5.3 Dynamic Range Reduction Acceleration

The tone mapping operators (TMO) for dynamic range reduction has been rapidly improved during last decade. Currently the topic is rather exhausted and next wave of improvements will be supposedly related to new research findings in the field of physiology and psycho-physic perception. Still the count of available methods is high. A brief introduction to this area is in Section 4 Methods for Luminance Scale Displaying.

One of the most complex physiologically influenced methods is [5]. This method (and many others, e.g. [6]) uses the bilateral filter for computing of the light adaptation. The filter is a bottle-neck in fast image processing. Though several attempts were made to accelerate the filtering, in 2011 we designed an approximation method with very small error and fastest computation so far. The method [7] is described in this Section.

Bilateral filtering is a nonlinear filtering computed as a weighted average of each pixel's surrounding. The weight is based on the spatial distance and the intensity difference. In most cases, the maximum weight is centered at zero differences of position and intensity.

$$B(p)=\frac{\sum_{s\in\mathbb{Z}^2} G_{\sigma s}(s)\cdot G_{\sigma i}(I_p-I_{p+s})\cdot I_{p+s}}{\sum_{s\in\mathbb{Z}^2} G_{\sigma s}(s)\cdot G_{\sigma i}(I_p-I_{p+s})} \tag{5.10}$$

The input image I_p and filtered image $B(p)$ are considered to be 2D discrete signals. The most used function for expressing the spatial and intensity scale weight functions are Gaussians: $G_{\sigma s}$ and $G_{\sigma i}$. The overall weight function is a product of both values.

Proposed Accelerated Algorithm

Computation of a bilateral filter is extremely slow. The aim is to design a very fast acceleration method, keeping the error below a recognizable level of the human visual system.

The bilateral filter used for HDR imaging uses broad spatial core and narrow stripe intensity core which increases the processing time even more. Taking advantage of this, we need to greatly simplify the spatial weight function and preserve the intensity core thin and accurate enough.

Unlike in most other attempts to accelerate the filter, the image is split spatially. In the following text, the presented algorithm will be explained step by step:

1. The image is split into tiles. Two different histograms are computed for each tile: histogram of the pixel intensity values and the same histogram where each count is multiplied by the intensity.
2. The histograms are convolved with a function close to intensity domain Gaussian $G_{\sigma i}$.
3. A spatial filter close to convolution with a space domain Gaussian $G_{\sigma s}$ is applied to the histograms. It means that the signal value is spread among the histograms in space, but not across each of the histograms.

4. The result image value is computed as the two histograms value ratio. An interpolation has to be applied.

Splitting the Image

The image is split into regular size, square tiles. Choosing an optimal tile size will be discussed below in the part Results. Using just equal splitting is not critical here. The rectangular tiles remaining at the far side of the image are treated in the same way as square tiles. The differential images (Figure 3 c,d) show no significant error along the right or bottom edge.

Now let us assume that the tiles are small enough and that the spatial weight $G_{\sigma s}$ can be approximated by a constant.

- Let θ be the tile size.
- Let $\varepsilon(p)$ be the set of pixels in the tile around pixel p.

Then the filter can be approximated as follows:

$$\begin{aligned} B(p) &= \frac{\sum_{s\in\mathbb{Z}^2} G_{\sigma s}(s)\cdot G_{\sigma i}(I_p - I_{p+s})\cdot I_{p+s}}{\sum_{s\in\mathbb{Z}^2} G_{\sigma s}(s)\cdot G_{\sigma i}(I_p - I_{p+s})} \\ &\approx \frac{\sum_{t\in\mathbb{Z}^2}[G_{\sigma s}(t\cdot\theta)\cdot\sum_{s\in\epsilon(t\cdot\theta)} G_{\sigma i}(I_p - I_{p+s})\cdot I_{p+s}]}{\sum_{t\in\mathbb{Z}^2}[G_{\sigma s}(t\cdot\theta)\cdot\sum_{s\in\epsilon(t\cdot\theta)} G_{\sigma i}(I_p - I_{p+s})]} \end{aligned} \tag{5.11}$$

Let $D(\varphi, p)$ be the denominator inner sum (sum through one tile pixels).

$$D(\varphi, p) = \sum_{s\in\varphi} G_{\sigma i}(I_p - I_{p+s}) \tag{5.12}$$

Let $H_\varphi(i)$ be the histogram of the set φ. Then $D(\varphi)$ can be stated as…

$$D(\varphi, p) = \sum_{i\in\mathbb{Z}} H_\varphi(i)\cdot G_{\sigma i}(I_p - i) \tag{5.13}$$

…which can clearly be expressed as one item of the sequence calculated as the convolution.

$$D(\varphi, p) = [H_\varphi(i) * G_{\sigma i}(i)]_{I_p} \tag{5.14}$$

The numerator sum through tile pixels $N(\varphi, p)$ can be simplified similarly. Let us define modified histogram $\acute{H}_\varphi(i) = H_\varphi(i)\cdot i$.

$$N(\varphi, p) = \sum_{s\in\varphi} G_{\sigma i}(I_p - I_{p+s})\cdot I_{p+s} \tag{5.15}$$

$$N(\varphi, p) = \sum_{i\in\mathbb{Z}} \acute{H}_\varphi(i)\cdot G_{\sigma i}(I_p - i) \tag{5.16}$$

$$N(\varphi, p) = [\acute{H}_{\varphi}(i) * G_{\sigma i}(i)]_{I_p} \tag{5.17}$$

Both inner sums across all pixels in one tile D and N are dependent on the tile histogram and on the source pixel intensity I_p. So one tile D and N can be pre-computed for all source pixel intensity values. To do so, the histogram of the tile has to be gathered and convolved with Gaussian $G_{\sigma i}$ which is described below.

Filtering the Histograms

We proposed the method for histogram convolution in [31]. It consists of the following steps:

1. The histogram is gathered already subsampled. The contribution of each pixel is distributed into the appropriate (closest) histogram values using a distribution function.
2. The sampled histogram is filtered by a set of exponential-average filters. This set of filters closely approximates convolution with a Gaussian, but is much faster.
3. The result value at any position is calculated using an interpolation function working with appropriate (closest) histogram values.

The error of the whole sequence can be enumerated.

The convolution of a function $f()$ with an exponential curve $(1 - \alpha)^n$ (EMA, Exponential Moving Average) can be computed within a single loop:

$$EMA_{ialpha}(n, f) = \sum_{m=-\infty}^{n} (1 - ialpha)^{n-m} \cdot f(m) \tag{5.18}$$

$$EMA_{ialpha}(n, f) = f(n) + (1 - ialpha) \cdot EMA_{ialpha}(n-1, f) \tag{5.19}$$

EMA2 combines two symmetrically flipped exponential curves and is performed by two loops, forward and backward.

$$EMA2_{\alpha}(n, f) = \sum_{m=-\infty}^{n} (1-\alpha)^{n-m} \cdot f(m) + \sum_{m=0}^{n} (1-\alpha)^{n+m} \cdot f(m+1) \tag{5.20}$$

Three of the *EMA2* pairs are combined (6 different exponential curves are used) to another proposed method called *3EMA*, a discrete convolution with an approximated Gaussian.

$$3\mathrm{EMA}(n, f) = 3.9 \cdot EMA2_{0.150}(n, f) - 3.9 \cdot EMA2_{0.247}(n, f) + 1.0 \cdot EMA2_{0.387}(n, f) \tag{5.21}$$

The above values were computed by using the simple error minimization. More details about the histogram filtering method are described in [31].

The sigma of the simulated Gaussian is altered through changing histogram bins count. The *3EMA* method simulates convolution with a Gaussian whose sigma is equal to 10 samples. So the number of histogram bins $|H|$ per function range Δf is calculated as

$$|H| = 10 \cdot \frac{\Delta f}{\sigma} \tag{5.22}$$

Let us point some other implementation details:

1. When the values are read from the convolved histogram, care must be taken to add/subtract 0.5 bin size to/from the index. The reason is that the *EMA2* method does shift the signal by a half bin to the right or left (depending on the particular implementation)
2. For the above reason and in order to avoid problems with linear interpolation, the histogram range should be expanded to both sides behind the maximum/minimum intensity by at least one bin.

Spatial Filtering

Let us assume both smoothed histograms N and D for each tile in the source image. The filter approximation would be as follows from (5.11).

$$\begin{aligned} B(p) &= \frac{\sum_{s\in\mathbb{Z}^2} G_{\sigma s}(s)\cdot G_{\sigma i}(I_p - I_{p+s})\cdot I_{p+s}}{\sum_{s\in\mathbb{Z}^2} G_{\sigma s}(s)\cdot G_{\sigma i}(I_p - I_{p+s})} \\ &\approx \frac{\sum_{t\in\mathbb{Z}^2}[G_{\sigma s}(t\cdot\theta)\cdot N(\varepsilon(t\cdot\theta), p)]}{\sum_{t\in\mathbb{Z}^2}[G_{\sigma s}(t\cdot\theta)\cdot D(\varepsilon(t\cdot\theta), p)]} \end{aligned} \tag{5.23}$$

Only one value from both smoothed histograms is taken for each tile. The tile contribution has to be weighted by a spatial Gaussian. In this case, however, the weighting of all tiles can be replaced by filtering the histograms in the spatial domain. It means that n-th bin of the histogram is filtered only with n-th bins of surrounding histograms. This 2D smoothing may be achieved by a separable filter. In our method only a small 3×3 convolution filter is applied, using the 3-item convolution horizontally and vertically. The size of the tiles is dependent on the maximum tolerable error. In the practice, their size is similar to the spatial Gaussian sigma (σs). In such a case, the spatial weight drops very fast. If the tolerable error is very small, the tiles become smaller, the number of pixels to the number of histogram bins ratio would drop and the whole method would become ineffective.

The spatial filtering can be even degraded by using only one spatial sample while the method is still usable. Both variants are compared further below.

Interpolation

The histogram being used for computing the pixel response should not be replaced instantly by it's neighbor between two neighboring tiles. Therefore, an interpolation should be applied to the histograms. The histogram used for a specific pixel response would be a weighted histogram of the pixel surrounding, where the weight centre is at the pixel. It makes the response closer to the original filter. In addition, the spatial interpolation avoids potentially visible edges at the tile borders.

According to the bilateral filter equation, interpolation should be applied to D and N separately. The simple linear interpolation would be as follows:

$$B(p) \approx \frac{(1-x)\cdot N(\varepsilon(t\cdot\theta), p) + x\cdot N(\varepsilon(t+1\cdot\theta), p)}{(1-x)\cdot D(\varepsilon(t\cdot\theta), p) + x\cdot D(\varepsilon(t+1\cdot\theta), p)} \tag{5.24}$$

x is the relative pixel position from 0 to 1 between two tiles. The bilinear interpolation for two dimensions is straightforward. Another spatial interpolating algorithm, Lánczos filter, has been implemented, but the result quality has not outperformed bilinear filter.

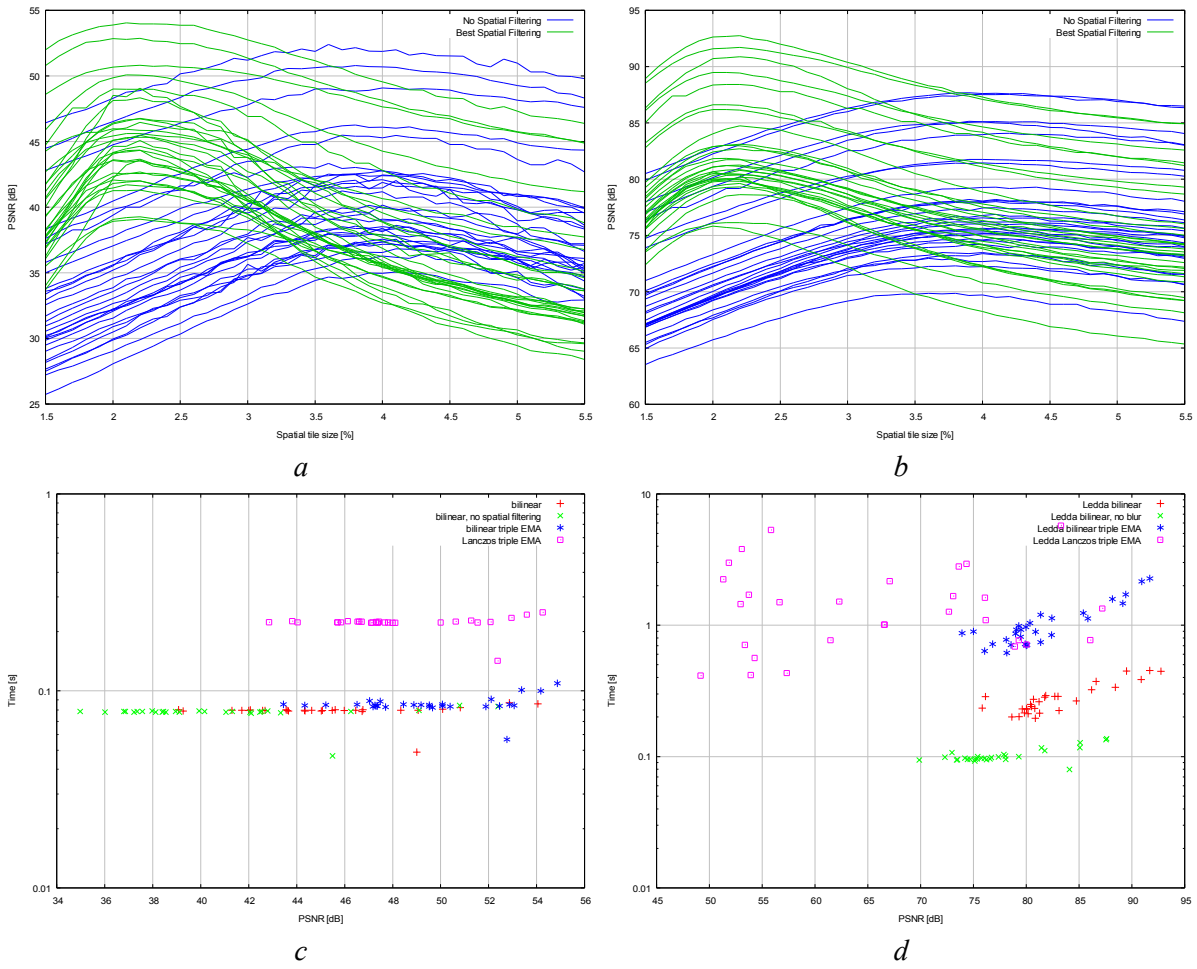


Figure 23: (a), (b) - spatial filtering performance; (c), (d) - various methods of histogram and spatial filtering performance for wide and narrow sigma configuration. Wide sigma (Durand) is used in (a) and (c), Narrow sigma (Ledda) in (b) and (d).

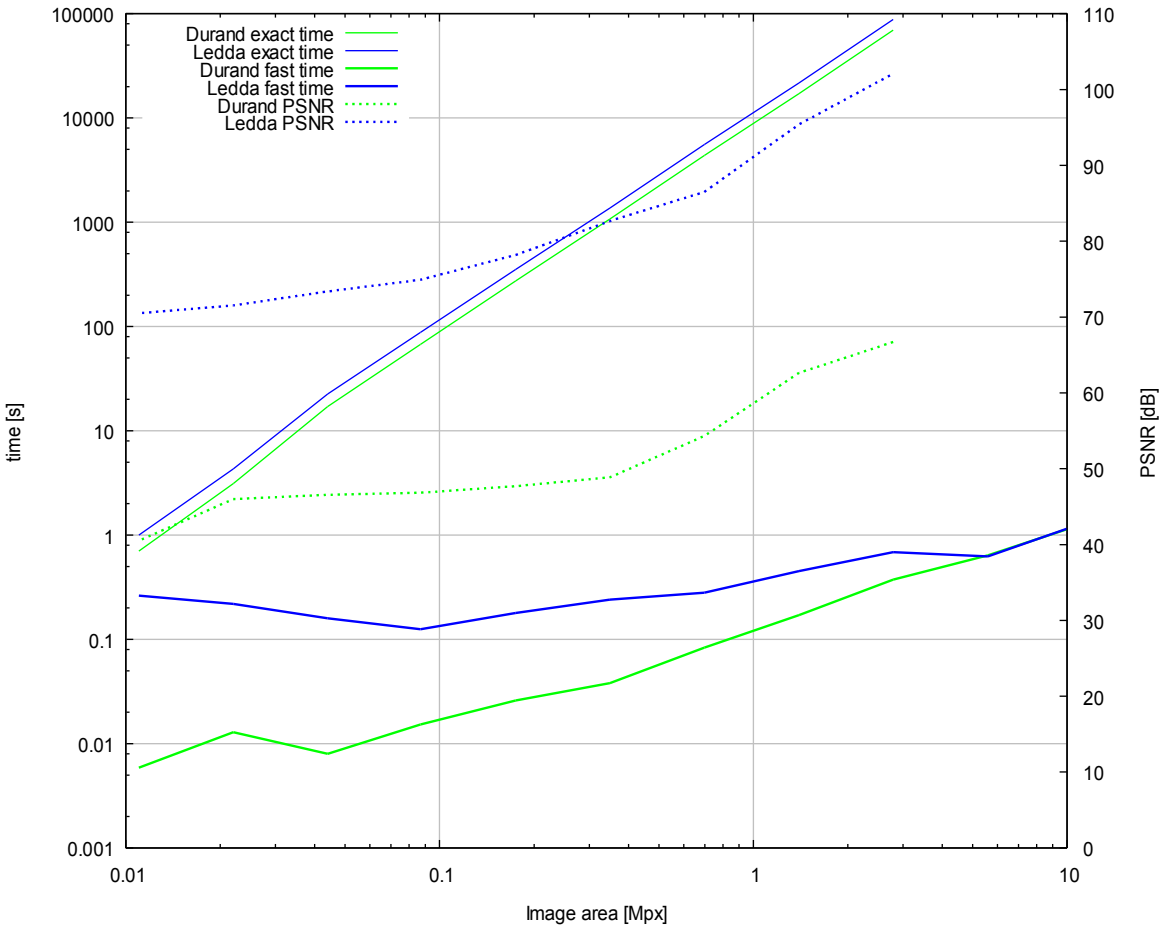


Figure 24: Dependency of a computation time and PSNR on the image area. Time of the exact bilateral filter computation (exact), time of the accelerated algorithm (fast) –Durand, triple EMA, bilinear tile filtering and Ledda, single EMA, bilinear tile filtering and PSNR for both filter settings.

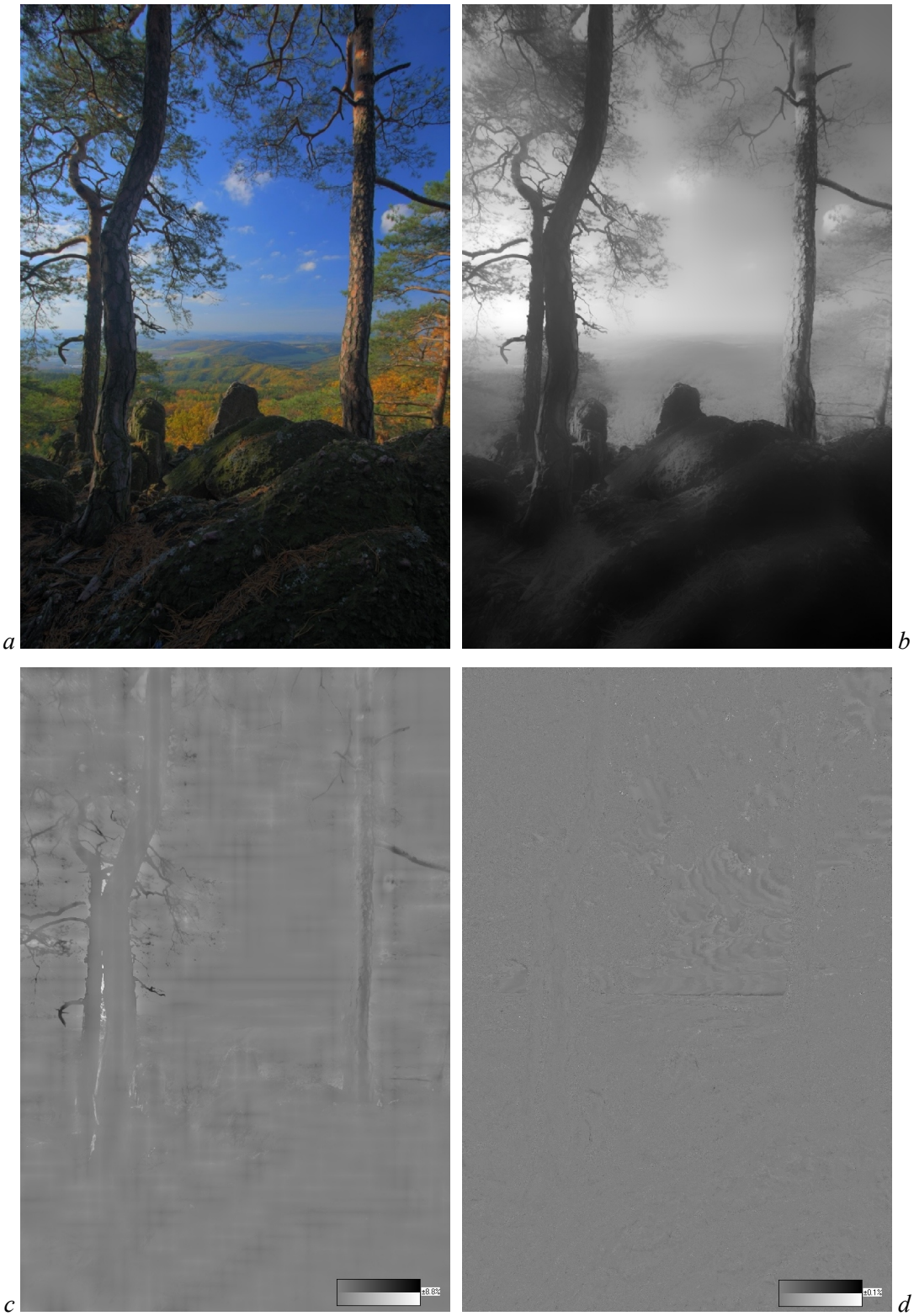


Figure 25: (a) Tone-mapped input image, (b) filtered image (Durand's σ), (c) differential image (Durand's σ, triple EMA), and (d) differential image (Ledda's σ, single EMA) (filtered image with Ledda's narrow σ is not visually different from the original image and therefore is not shown).

Results

For testing, the real bilateral filter has been implemented. The only restriction is spatial radius, which is limited to 5σ where the weight is about $3.7 \cdot 10^{-6}$.

Twenty-nine (29) images of different but ordinary topics have been filtered. The area of each image was approximately 0.7 megapixels. Two sets of tests have been run, approximating two state-of-the-art papers [5, 6] on a bilateral filter adjustment. For each test, the image was also filtered by using the exact bilateral filter. The PSNR value has been computed for measuring the approximation precision as used in [32].

Durand and Dorsey prefer an intensity sigma $\sigma i = 0.4$ on a logarithmic image (log 10) which corresponds to $\sigma i = 4$dB. Ledda, Santos, Chalmers use a floating sigma value linearly dependent on the intensity $\sigma i = 0.15 \cdot i$. In our test, the floating sigma was approximated by a fixed $\sigma i = 0.6$dB on the logarithmic image (for both ideal model and accelerated filtering, so PSNR does not cover floating sigma approximation).

The filter spatial sigma was always 2% of the image's longer side.

Firstly, we tested the performance in PSNR against tile size. For each tile size we have found the best spatial filter values. Value Side Point Blur (*SPB*) was defined within a range $\langle 0; 1/3 \rangle$. The tile spatial smoothing is performed as described above, by convolution vertically and horizontally with a vector (*SPB*, 1-2· *SPB*, *SPB*). The results (see Figure 23 *a* and *b* show that optimal tile size is about 2.2% of the image size and this corresponds to 1.1σ. In the same graphs non-filtered method performance is plotted (blue curves). The non-filtered tile size should be about 4% of the image size, which corresponds to 2σ. PSNR is about 5dB below the filtered variant.

The next two graphs show the search for the best spatial filter values on the same set of images. The curves show the best Side Point Blur (the solution with the highest PSNR) for different tile sizes. 0.1% tile size and 0.02 Side Point Blur steps were used for searching. The graph shows, that for the local neighborhood, the tile size can be compensated for by spatial tile filtering. The dependency is close to linear. Figure 23 *c* and *d* show several tested methods.

Bilinear or Lánczos is the interpolating method among tile histograms (see the part Interpolation). Clearly, Lánczos interpolation consumes more computing time, but unexpectedly the PSNR is worse than with bilinear interpolation.

Skipping the spatial tile filtering loses some PSNR and spares some time in both sets as mentioned earlier (first the pair of graphs show it more precisely)

The last result is about using triple EMA histogram filtering. In the first set of tests (wide intensity Gaussian) triple EMA raises PSNR by about 4dB at about 10% more time cost. In the narrow intensity Gaussian tests however, the computing time rose several times without significant differences in PSNR and some results were even worse using triple EMA. But still the PSNR never dropped below 70dB using a single EMA with spatial tile filtering, which means that narrow intensity Gaussian are better to compute using a single EMA.

Another set of tests has been run for a different scaled image. As in the previous tests, both intensity sigma settings (Durand and Ledda) were tested against their exact computation (see Figure 24). For very large images (over 5 megapixels) the computing of the exact bilateral filter would take an unacceptably long time, so only the approximation time is shown. The time complexity order of the exact bilateral filtering is clearly quadratic. The complexity order of the accelerated algorithm is close to linear, yet some deceleration is apparent for smaller images. This might be caused by computing with too many histograms of very small areas. The algorithm would be obviously ineffective for very small images. Also, PSNR tends to rise with the image area.

The visual difference between fast and precise bilateral filter is shown in Figure 25.

5.4 Resampling

Resampling methods were described in the Section 3. But it has to be said, that ideal resampling is not necessarily best for display devices. The reasons are:

1. Display pixels are not ideal samples. Ideal sample would be close to Gaussian or Sinc function with very low frequency domain response above half of the sampling frequency.
2. In the HVS there is no lo-pass filter, which would have suppressed high frequency harmonic signal caused by inadequate samples. Or to be more specific, there is a filter suppressing high frequencies in HVS, but the inhibition is not very steep and it depends on the observation distance.

These reasons results from the physiology part (Section 2). The problem is examined in detail on the following pages.

Resampling and the Color Information

In practice, color information is mostly stored as radiance of three base colors. Total luminance is given by the three values. Alternatively the luminance channel is stored as one value apart from hue and saturation.

Both records are convertible and are capable of carrying the information processed by HVS. In fact, the retina records the signal as three base color values, but the pathways to visual cortex contain three different signals. Apart from the brightness, there are signals “red to green” and “blue to yellow” (see Section 2.3, part Color Perception). These values represent a Cartesian parameters of a polar plane given by hue and saturation (see Figure 26).

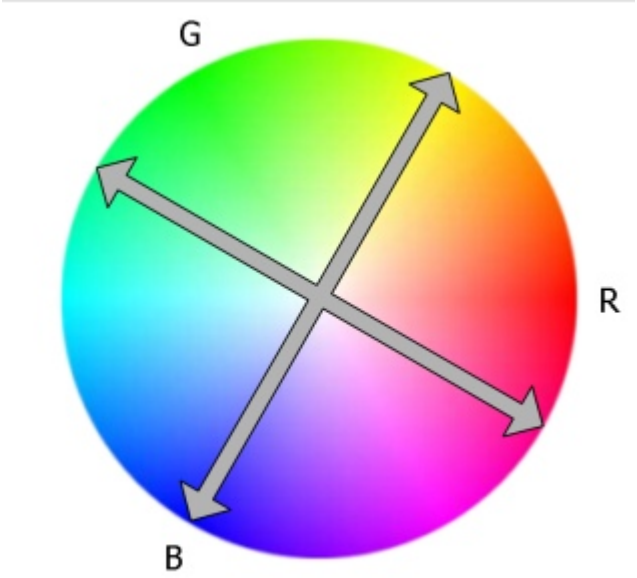


Figure 26: Hue circle and color opponency dualities in the HSV

If the “hue & saturation” is represented by base colors (RGB) or any Cartesian parameters, interpolation between colors can be performed by simple linear combination of the values.

The question is, how should be the color treated in resampling. Strictly speaking, the observed color can be composed of base colors. Mixing of the light sources obviously results in a linear combination of the base colors. If the resampling is performed by a linear filter, it could be calculated separately per each base colors. Yet two problems have to be taken into account:

The first problem is that in practice, digital image processing is often performed on nonlinear luminance values. Commonly used 8-bit quantization is very rough in linear luminance recording. Therefore, nonlinear gamma-correction is mostly applied to the values before quantization. This kind of data should be not filtered by linear filter, even if the data had represented only the luminance without color. And the error of the color is subjectively even more visible. If the result is to be correct, any nonlinear correction of the values being filtered is unacceptable.

The second problem comes with the resampling optimization. Measured characteristics of the primate visual system can be used to optimize the resampling. But these measurements [1, 2] are in colorless domain. The color information is separated in the HVS and we can not assume that the hue signal is processed in the same way. On the contrary, the hue visual acuity is approximately one third of the brightness acuity (see Section 2.3, part Color Perception). The source of this difference is unclear, yet it means that the hue information may be resampled without physiology influence. Before any more information about the hue pathways is available, the most correct way of resampling is the state-of-the-art Lánczos filter.

But the low color acuity also means, that enhancing of the brightness channels precision may be worth lower precision in the color channel. In most displays, pixels are compiled of separated units with different base colors, like horizontally aligned rectangles in most of modern LCD displays. Specialized resampling methods for spatially split base colors have been presented, e.g. [28].

Brightness Channel Resampling

Generally accepted approach in image resampling comes from the signal theory, described in Section 3. Basic assumption is sufficient information about the source image. In the ideal case, the image would be continuous. In practice, the image should be sampled with sampling frequency several times higher than the sampling frequency of the display. According to the Nyquist theorem, only harmonic signals up to the half of the sampling frequency may be stored in the sampled signal. Thence higher harmonics are to be filtered out.

Hi-Pass Filtering

Now, let us leave the exact ground for a while. When we search the area of practical computer graphics, photographers, DTP workers, editors etc., we are to find that the used methods do not follow the approved theory. A dominating opinion tells, that a sharpening effect has to be applied to a digital image after each resampling. The exact parameters of the effect are a subject of fierce discussions. This statement, is not supported by any serious research, and in fact it can be easily disproved, as will be shown further below. Nevertheless, the general view can not be underestimated. Aside from untrustworthy discussion forums, the sharpening is advised by some respected authorities and Google finds nearly one million pages with the keyword “Unsharp mask”, Adobe Photoshop version of said filter.

Before contradicting the sharpening approach, let us consider, if and why there could be any sense in it.

Most of these effects are simple hi-pas filters, often realized by subtracting a lo-passed and scaled original image (an open-source image manipulation program Gimp implements “Unsharp mask”, too, so Gimp's version of the source code is available). The recommendations tell that optimal effect is achieved by very small spatial kernel and strong weight. These parameters form the simplest hi-pass convolution filter with kernel size 3 (Figure 27).

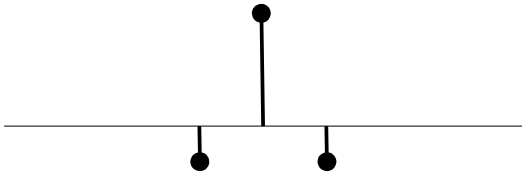


Figure 27: FIR hi-pass filter, size 3 samples

This filter strongly enhances harmonic signals close to the Nyquist criteria, one half of the sampling frequency.

When a typical user is observing a common display, the half-sampling frequency is below the HVS limits, but sure beyond the maximal sensitivity (see Figure 28). It follows that the sharpening effect clearly can make some component of the image easier recognizable.

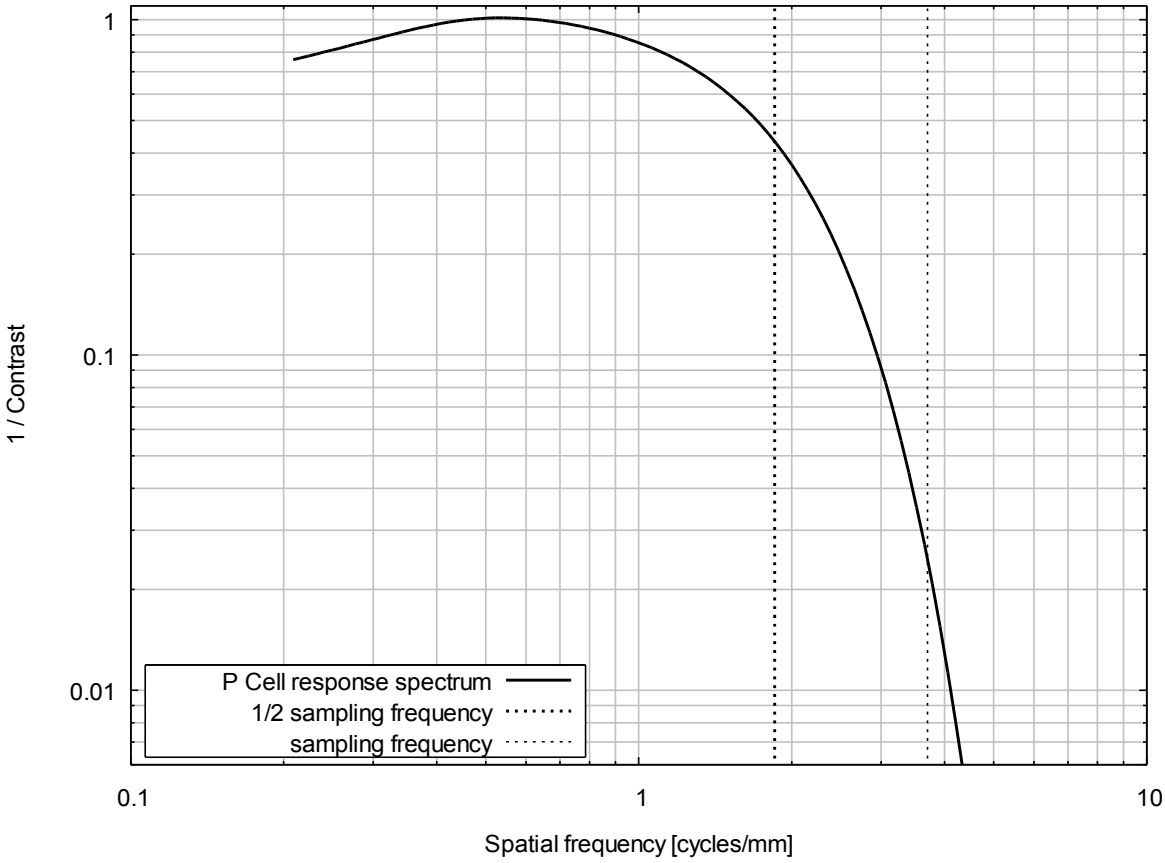


Figure 28: Spectrum of scaled P ganglion cell spatial response from [1] projected to the display for a typical observer. Measurement of the visual acuity, observation distance and computing of the response is explained below in Sections 5.5 and 5.6. Sampling frequency markings are plotted for the used display with pixel pitch 0.270mm.

So far, it makes sense that the sharpening may seem to add more details into the resampled image. But it doesn't prove, that the result after sharpening would be any closer to the original image. From the technical point of view, the approach has one strong weakness: It tries to enhance the harmonic components, which were mostly suppressed during the resampling. Even if the sharpening influence to the perception quality was admitted, amplifying a quantized signal, which has been just strongly suppressed is completely wrong on principle. In other words, the two steps used in sequence amplify

quantization noise. So even if there was a reason not to suppress some harmonic components of the image, the filter should have be applied in one step, nos sequentially.

Pixel Shape Problem

In the signal theory, the ideal sampling is supposed in most cases. But this does not apply to the real display devices. The spatial characteristic of the pixel emittance is close to rectangular, width of the pixel is approximately between 80% and 90% of the pixel pitch (see Figure 29)

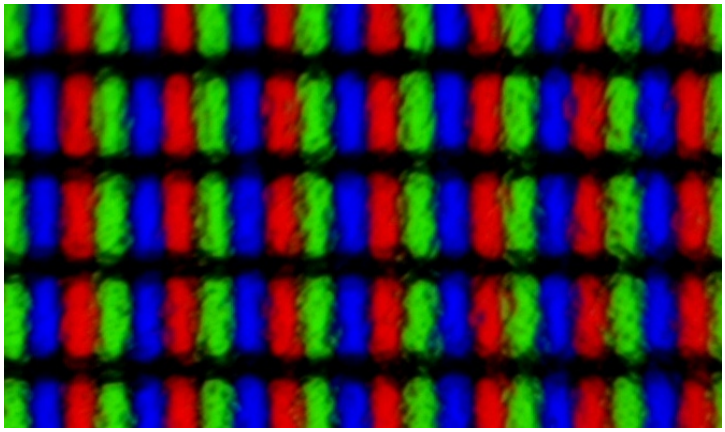


Figure 29: Detail of LCD screen, pixel pitch 0.28mm

The statement, that signal can not contain harmonic components of higher frequency than one half of the sampling frequency doesn't apply here. The high frequency harmonics above the sampling theorem are contained in the pixel shape.

Let us show an example. A one-dimensional signal consists only of one steep edge (Figure 30). The signal is filtered by an ideal lo-pass filter and sampled to the screen. The edge is positioned at the middle of a pixel (a). The sampled signal would look like two identical steps (b). The sharpening effect broadens the differences between the neighboring pixels (c). So far, we can still believe, that once the image was processed by HVS, the result could be closer to the original edge. But the situation becomes different, when the edge is positioned exactly between two pixels (d). In this case the filtered and sampled almost exactly copies the original signal (e). The sharpening effect obviously damages the signal (f).

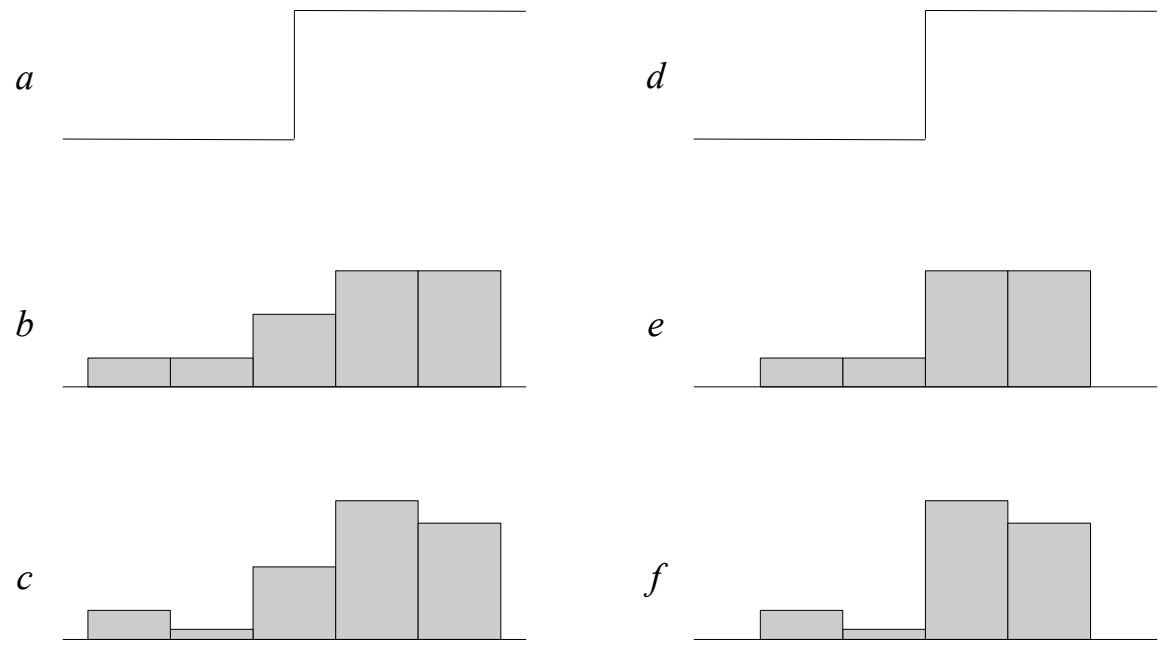


Figure 30: Resampling and hi-pass filtering of an ideal step function

This example demonstrates two points. Firstly, the popular sharpening is unmistakeably wrong. And secondly, the pixel shape affects the result and must be taken into account, when the perception is optimized.

Resampling Optimization

On the following pages, a resampling optimization approach is presented. It combines pixel shape, HVS spatial response and measurement of the optimal observation distance. The goal is optimal perception, which means optimal transfer of the whole system, including the physiological part (Figure 31).

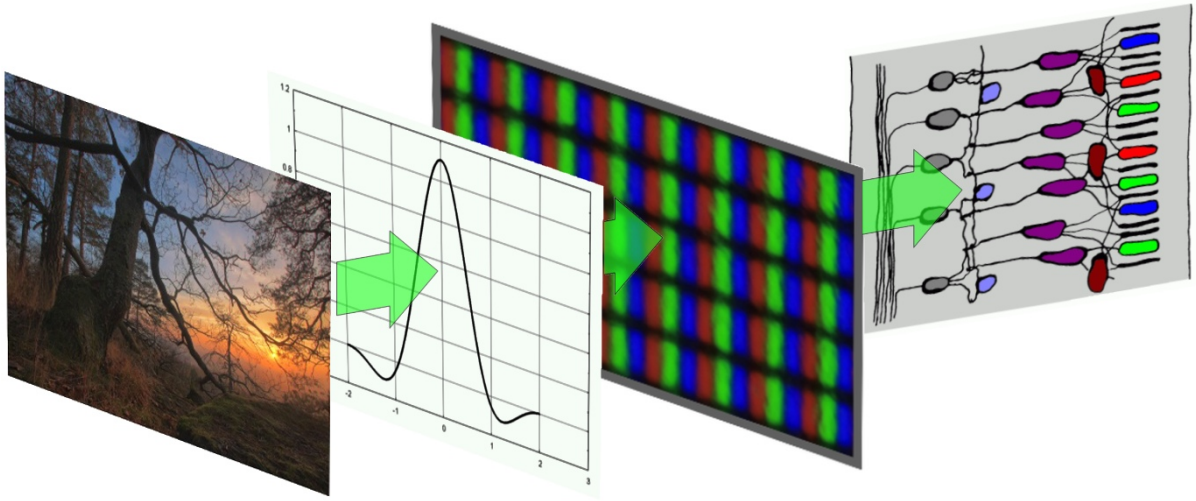


Figure 31: Perception of an image on a display: Image is filtered and sampled, rendered via the display pixels and processed by the HSV

Most displays render color information using spatially split pixels into three rectangular areas emitting different base colors. Methods for this devices has been already presented, e.g. [28]. For this reason, and because of easier explanation, we present a one dimensional resampling optimization, which can be used vertically on these displays. 2D variant of our method can be derived straightforward and used for any display device with uniform rectangular pixels.

5.5 Relation Between Visual Acuity and Optimal Observation Distance

Users tend to view the display from the so-called "comfortable distance". However, the term "comfortable distance" is very vague, so it needs to be narrowed in order to be more specific. The aim is to find an ideal observation distance which is the best one for examination of details on a still photograph rendered via the display device. As the "examination of details" is still not specific enough and can even be different for different users, the task can be made more specific through description and comparison of the achieved quality of image perception. Each rendered picture was processed in several different ways and the users were to subjectively choose which version they like. This lead into spontaneous localization of the users in the optimal observation distance suitable for comparison of the image details.

The question is how does the comfortable observation distance correspond to the visual acuity. Unlike the standard visual acuity measurement, the optotypes were placed at the chosen distance, onto the display surface. The real visual acuity value is, therefore, relative to the chosen distance, which was measured as well. This approach ensures that the conditions, mainly the focal plane of the eyes, are close to those that apply during observation of the display. In some subjects, the different distances would also cause a need to change glasses which also means that the results in the subject with and without the glasses might not correspond.

23 subjects, 6 women and 17 men of ages between 23 and 86 years were tested. Each of the subjects was seated in front of a display on a movable chair and also instructed to move freely and position itself into a comfortable position. Several images were then shown in a sequence. Each of the images

was filtered by four slightly different hi-pass filters. All four variants of the same image were then rendered on the screen side by side. The order of the filters was always random. The subject was asked to rate all of the four variants by selecting “like” “dislike” or “neutral” icon. The testing screen is shown in Figure 32. The differences between the digital filters were very tiny; therefore, the observer chooses the best conditions for careful and detailed examination of the images. Most persons stopped moving forward and back after first two image sets and stayed at the position for the rest of images. After the image rating was finished, they were asked not to move. The screen was then turned white and overplaced by the visual acuity measurement chart (see Figure 33). Subjects were asked to select the finest resolvable pattern of horizontal stripes. The distance eye to display was measured during the test, without disturbing the subjects (using triangulation). The optional reading aids were used throughout the test.

Two persons have chosen very long observation distance and were not able to recognize even the roughest optotype (3.35 cycles/mm), possibly because of being affected by an initial phase of hyperopia with no correction. The acuity could not be measured, therefore the subjects were excluded from the statistics.

In the first test, many subjects reported difficulties in optotypes resolving, because with some of the densities, the pattern was visible at the field border only. To prevent the inaccuracy caused by this fact, another set of patterns were prepared with faded borders (Figure 34) and the subjects acuity was re-evaluated after the tests were repeated.

Important data

Display model	HP LP2465
White absolute luminance	60.8 $cd \cdot m^{-2}$
Black absolute luminance	0.245 $cd \cdot m^{-2}$
Luminance range	1:248
Pixel spacing	0.270mm
Display was calibrated with sensor	(Datacolor Spyder 3 Pro)
Gamma correction	2.2
White balance	6500K
Diffuser underneath the optotypes	
Thickness	3mm
Luminance loss	0.673 (-1.7dB)
Optotypes	
Seven striped optotypes optically transferred to inverse film	(Fuji Provia 100F RDP3)
Minimal density	3.35 $cycles \cdot mm^{-1}$
Maximal density	4.78 $cycles \cdot mm^{-1}$
White absolute luminance	15.1 $cd \cdot m^{-2}$
Black absolute luminance	0.202 $cd \cdot m^{-2}$
Contrast	1:74
Testing room	
Maximal illuminance measured at the desk	65 lx

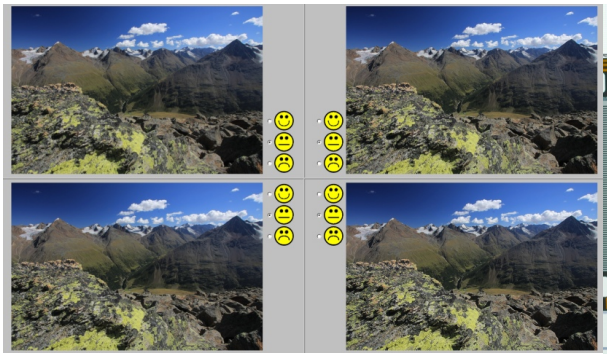


Figure 32: Testing screen

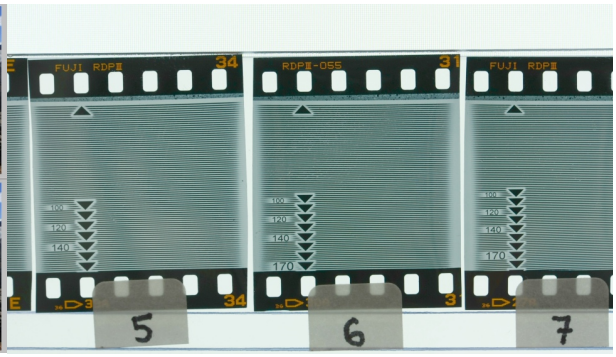


Figure 33: Optotypes

Figure 34: Optotype microscopy scan

Figure 35: Optotype marks for density measuring

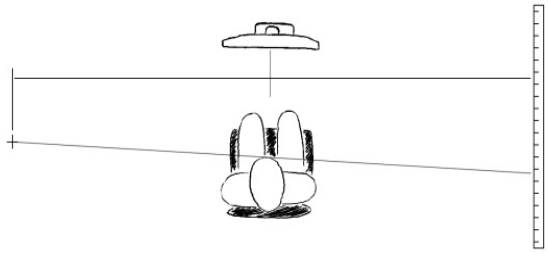


Figure 36: Distance measuring

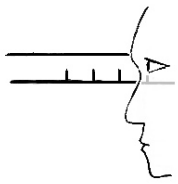


Figure 37: Distance measuring

The distance from eye to the display was read from a marked point on a scale placed far on the side, parallel to the display axis (see Figure 36). The parallax error was compensated by proper scaling of the meter. The reading included estimation of the distance covered by subjects nose (approximately 1-2 cm, see Figure 37). Note that millimeter precision was anyway unable to reach due to permanent micro-movement of human body. Head fixation on the other hand would spoil the comfortable position choice.

The optotypes with exact density could not be made without specialized optics. The optotype set was made with approximate scaling. Control marks were printed among the stripes (see Figure 35).

For each of the optotypes the distance between the marks was measured with precise caliper and the density was calculated from the scale factor. Therefore the optotype set do not follow linear nor logarithmic series.

Angular acuity was computed from the relative acuity and eye-to-optotype distance as follows:

$$A_a = 10 \cdot A_r \cdot d \cdot \frac{\pi}{180} \tag{5.25}$$

where A_a is the angular acuity in cycles per degree, A_r is the relative acuity in cycles per millimeter and d is the distance in centimeters. The 3mm thickness of the diffusing plate underneath the optotypes was omitted in this calculation because of the distance precision issue mentioned above.

Results

Several subjects were complaining about the difficult judgment and tiny differences, yet 20 of 21 subjects showed clear preference in some methods. The order of the image processing methods applied on the images to be observed was random for each person, so it is clear that the image differences were possible to recognize by the subjects. One person with almost stochastic preference among the methods was excluded from the statistics, because of possible misunderstanding of the instructions and choosing a far position where no differences might be recognizable.

We supposed that the results may be biased by subjects with possible professional interest in the field of digital image processing or photography. For this reasons, people were asked to state such level of interest in image processing and specifically marked in the results. In the charts they (aside from "Men" and "Women") formed the third group "Graphics" without any further sex differentiation.

Both angular acuity and relative acuity are statistically non-symmetrical values. Even if the divergence is not high, acuity values should be by their nature symmetrical in logarithmic scale. Therefore, mean values and standard deviations were computed in $\log_{10}$.

Angular acuity	32.3 cycles·deg^{-1}
Ang. ac. deviation	0.118 $\log_{10}$ cycles·deg^{-1}
Relative acuity	3.98 cycles·mm^{-1}
Rel. ac. deviation	0.024 $\log_{10}$ cycles·mm^{-1}

Although the acuity varies, it shows strong correlation with the preferred observing distance (Figure 38 and Figure 39). The results show that the relative spatial acuity in preferred distance has much smaller deviation than the angular acuity.

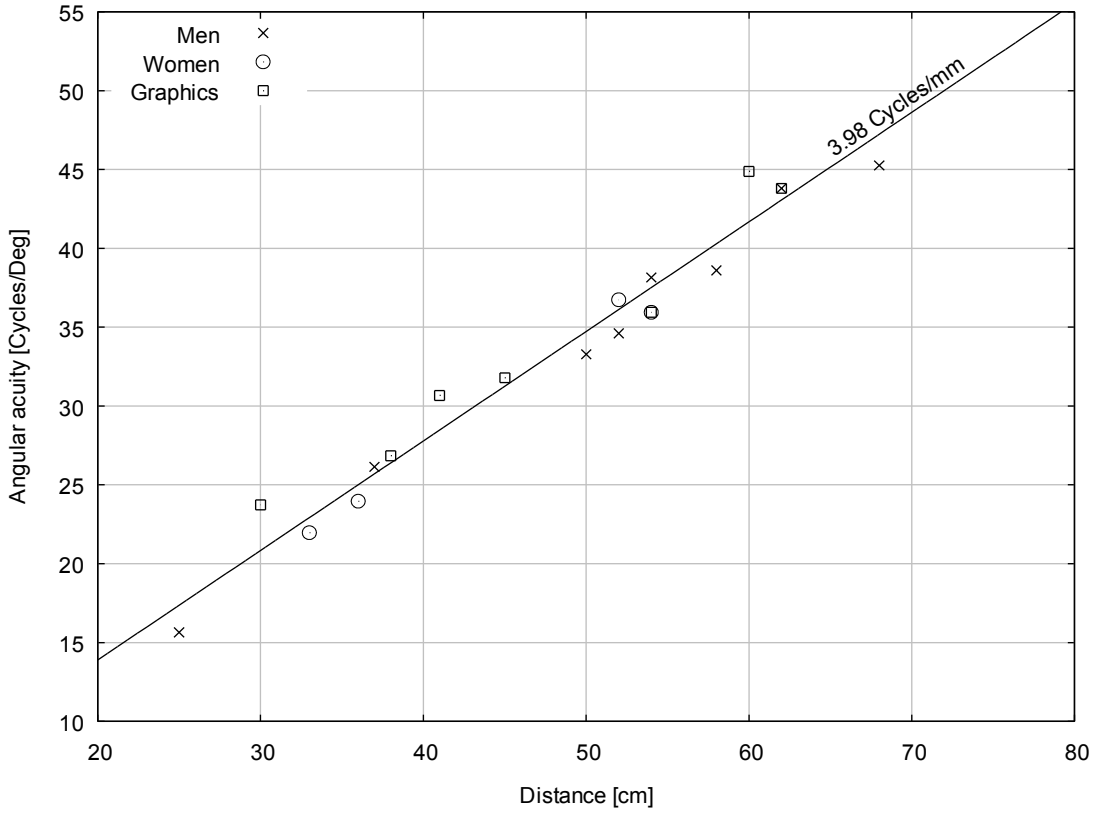


Figure 38: Correlation between angular acuity and preferred observing distance

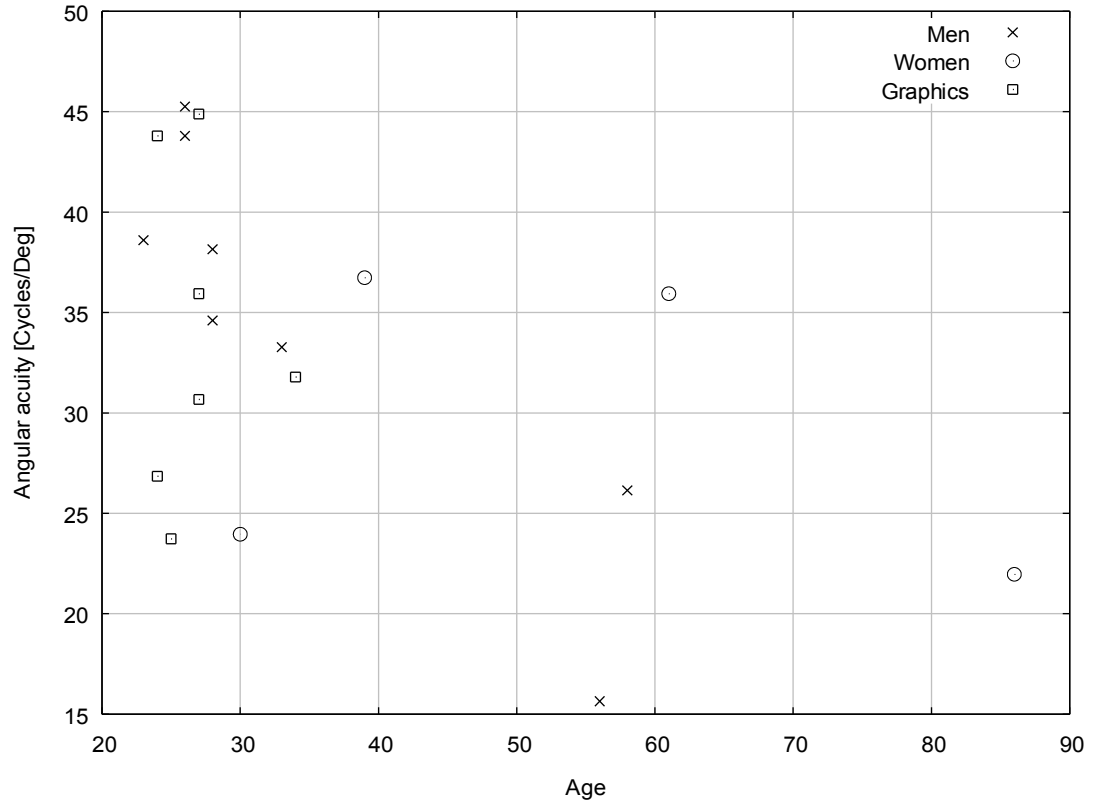


Figure 39: Distribution of angular acuity across the age

The fact, that human observers tend to normalize acuity at the display level opens new possibilities in image processing. It follows that the relative spatial acuity could be possibly applied with the HVS spatial response.

The test could be also extended by measurement with higher density displays. Such devices were so far not included for couple of reasons. Partly, because of poor availability at the time when the project started. However, the main problem was that many devices in consideration had either no anti-glare surfacing or extreme luminance instability with vertical observing angle. In both cases, it was impossible to prepare stable measurement conditions.

It is not clear, if the divergence of the relative acuity would stay so low with high density displays. The comfortable observation distance is surely limited. Most likely, the unity among all observers would spread according to individual shortest focal length or even personal preference.

5.6 Resampling Filter Optimization

Choosing the Retina Spatial Response

The spatial response of many different cell types has been measured, mainly in cat, fish and monkey, see Section 2.2, part Retina Cells Receptive Fields. Here, the most important is to get close to the conditions we are interested in.

1. The retina structure differs a lot among the various genus, therefore the recording in primates is the most accurate.
2. The aim is to refine the visual perception. It means, that the more of the visual system can be modeled, the better. The most interesting would be the recording of the ganglion cells, the last cells to process the signal before sending it to the brain, or recording of the brain cells themselves.
3. The examining of fine details is performed by cones in the fovea. Central vision in the fovea provides the finest spatial resolution. The peripheral retina inhibits several times lower spatial frequency than frequency passed by central retina. If the image is optimized for the fovea and observed by peripheral retina, the difference in high frequency signal would be negligible.
4. Parvocellularly projecting (P) ganglion cells are more sensitive to spatial details. Pathways from P cells are therefore essential in optimization of spatial perception quality.
5. The spatial attributes of the receptive field are affected by the light level, so recording in the average luminance of the common monitors (approximately 10-100cd·m^{-2}) is preferred.

Data from [1] has been selected as the most appropriate measurement. The macaque monkey receptive field of P and M retina ganglion cells was measured. The mean luminance was 40cd·m^{-2}, see Section Retina Cells Receptive Fields, Figure 6.

The response has to be recalculated from section to integral along lateral domain. Figure 40 Shows the one-dimensional integral of the spatial response, Figure 41 shows comparison between human CSF [17] and the spatial response spectrum. The small difference in frequency is caused firstly by greater eccentricity of the ganglion cells compared to the size of fovea centralis. Possibly, the difference may be caused even by different genus, namely different posterior nodal distance.

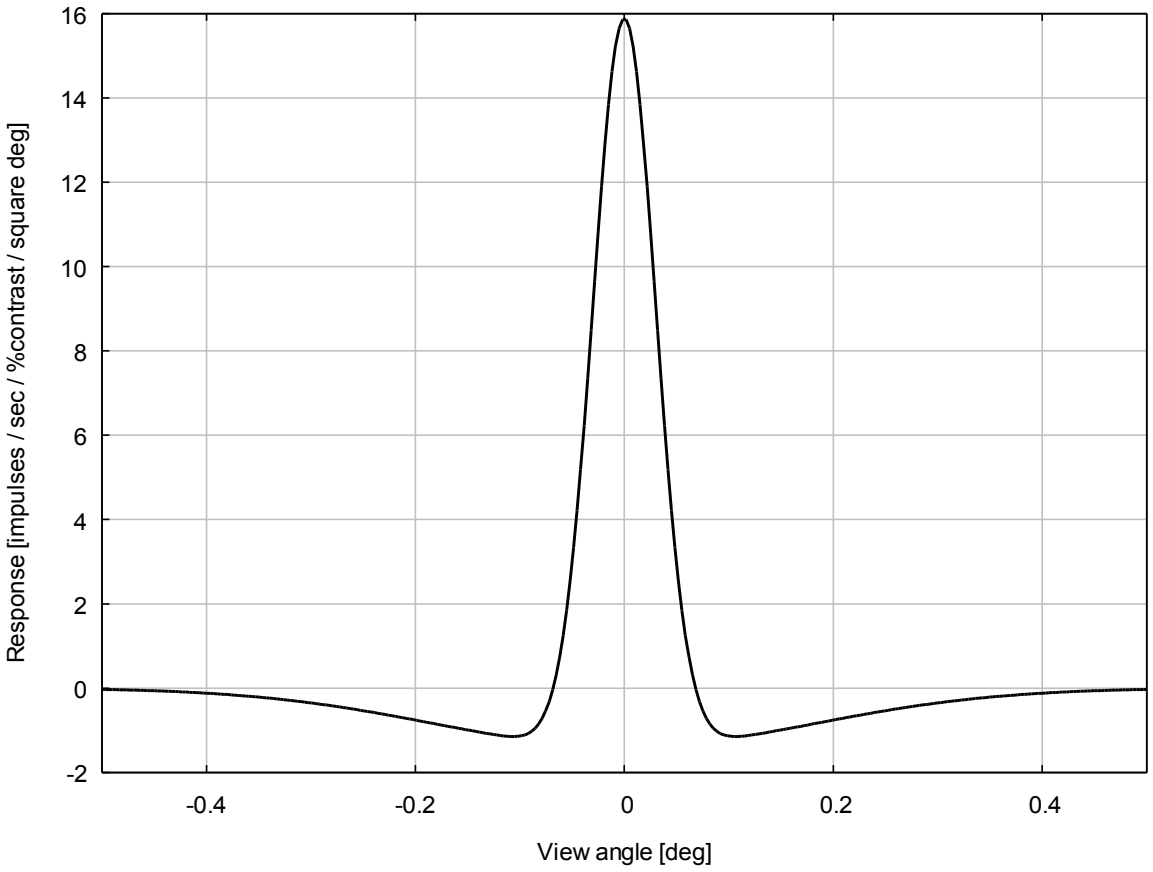


Figure 40: Small eccentricity P ganglion cell spatial response [1] recalculated from section to one-dimensional integral.

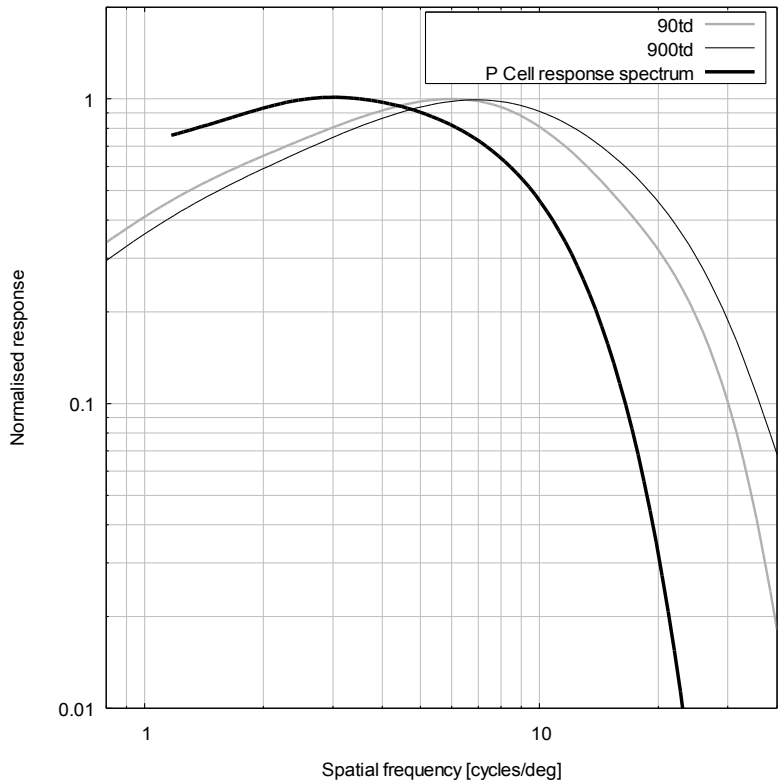


Figure 41: Human CSF [17] compared with macaque P ganglion cells spatial response spectrum [1]. All functions are normalized.

Relative Acuity

People with worse acuity prefer shorter observation distance and vice versa. The measurement described in previous Section showed that people across very different angular acuity tend to chose a distance, where the relative acuity is almost constant.
The low divergence of the relative acuity allows for using a uniform response of the visual system for any observer. This approximation causes an error evaluated as follows.

Let us assume, that a lower visual acuity is caused mainly by optical aberration in the eye's optical part. The aberration can be expressed as a convolution of the retina cell spatial response with a point spread function (PSF). A general aberration PSF is a Gaussian.

Changing of the observation distance results in linear scaling of the receptive field projection at the observation plane.

So the receptive field projection onto the display plane is given by convolution with a distorting Gaussian and linear scaling.

The angular acuity standard deviation is approximately 0.12 $\log_{10}$ cycles·deg^{-1} (see Section 5.5), which corresponds to acuity ratio:

$$R_A \approx 1.31 \tag{5.26}$$

If the angular acuity is R_A times worse and the relative acuity should stay the same, the person needs to shorten the observation distance by the same ratio.

The chosen spatial response was approximated as a sum of center positive and surround negative Gaussian. The ratio between center and surround sigma, is:

$$R_{CS} = \frac{\sigma_s}{\sigma_c} = \frac{0.18}{0.03} = 6 \tag{5.27}$$

Let us simplify the situation and assume, that R_A times worse visual acuity corresponds nearly to R_A times greater sigma of the receptive field center Gaussian (surround Gaussian spectrum is six times narrower so it's response to just resolvable frequency can be omitted). This change of acuity should correspond to convolution of the spatial response with a distorting Gaussian.

Convolution of two Gaussians with σ_1 and σ_2 is another Gaussian with:

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2} \tag{5.28}$$

Thence the person with R_A times worse angular acuity has spatial response with:

$$\begin{aligned} \sigma_c' &= \sigma_c \cdot R_A \\ \sigma_s' &= \sigma_s \frac{\sqrt{R_{CS}^2 + R_A^2 - 1}}{R_{CS}} \end{aligned} \tag{5.29}$$

If the spatial response distortion is approximated with scaling by R_A, the error of surround sigma would be:

$$E_{\sigma s} = \frac{R_{CS} \cdot R_A}{\sqrt{R_{CS}^2 + R_A^2 - 1}} \approx 1.299 \tag{5.30}$$

This is still acceptable error, because the surround sigma is six times greater. Note that the ratio between center and surround Gaussian area is not affected. The distorting Gaussian area is equal to 1 so it does not change the convolved functions area.

The receptive field scaling is given by it's spectrum. The measured relative acuity at the display plane corresponds to the spectral frequency with loss equivalent to the optotype contrast (see Figure 42).

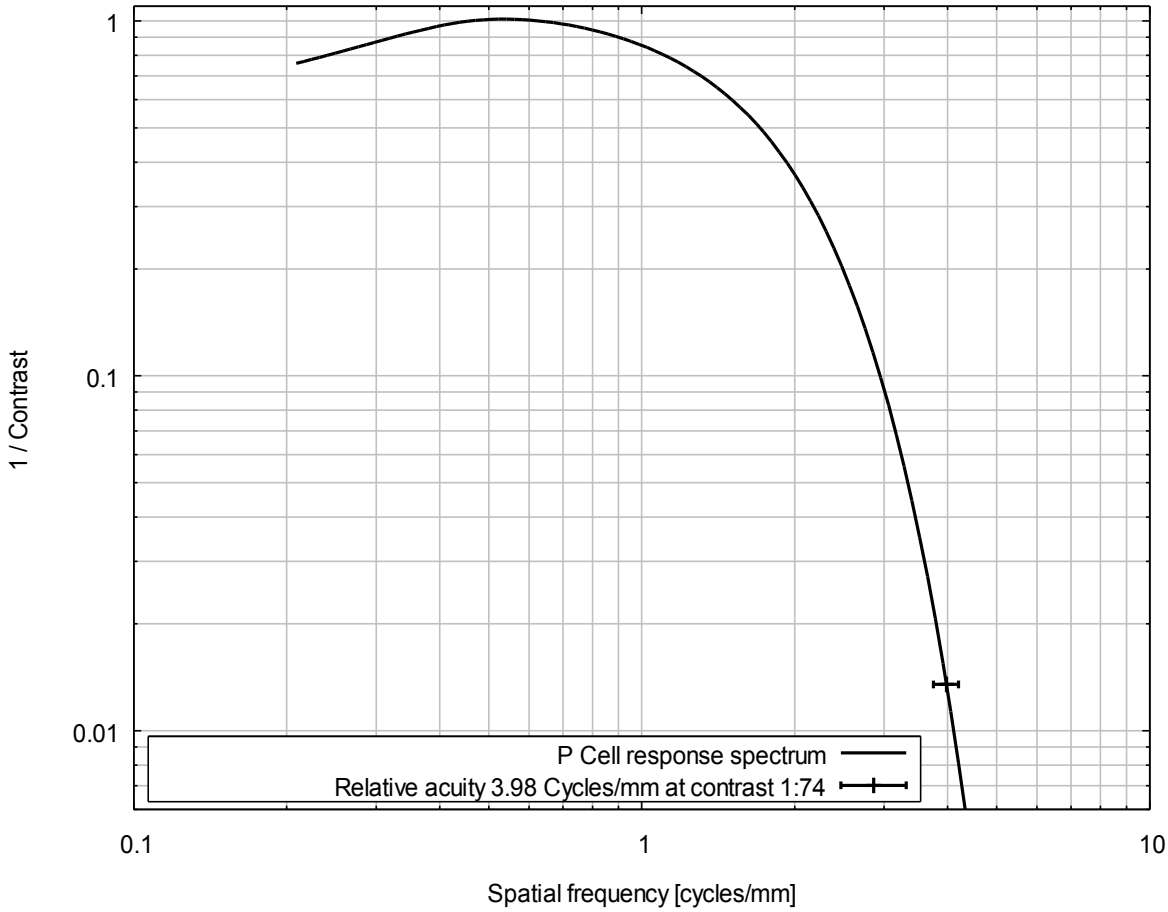


Figure 42: Spectrum of scaled P ganglion cell spatial response from [1]. The scaling corresponds to the measured relative acuity. Relative acuity with standard deviation is plotted at the optotype contrast level 1:74

The scaling is 7.570mm/deg. It gives the spatial response sigma values in display lateral domain:

$$\begin{aligned} \sigma_{Ds} &\approx 1.010\text{mm} \\ \sigma_{Dc} &\approx 0.168\text{mm} \end{aligned} \tag{5.31}$$

Filter Optimization

When the image is post-processed, rendered on a display and observed, the whole process can be described as follows.

$$Sens(x) = [Img * Filter \cdot \delta * Pixel * Eye](x) \tag{5.32}$$

The equation is in 1D space. Figure 43 illustrates the process. The image *Img* is firstly convolved by *Filter* and then sampled by multiplying with Dirac impulses δ. The samples are convolved with a spatial characteristic of a screen *Pixel*. As our method should be applicable to most displays, rectangular characteristic with 85% size of the pixel-to-pixel distance was assumed, however for particular display the characteristic is better to be measured. The convolution with *Pixel* gives the image on the screen. Finally it must be convolved with the properly scaled retina response *Eye* to get the sensation.

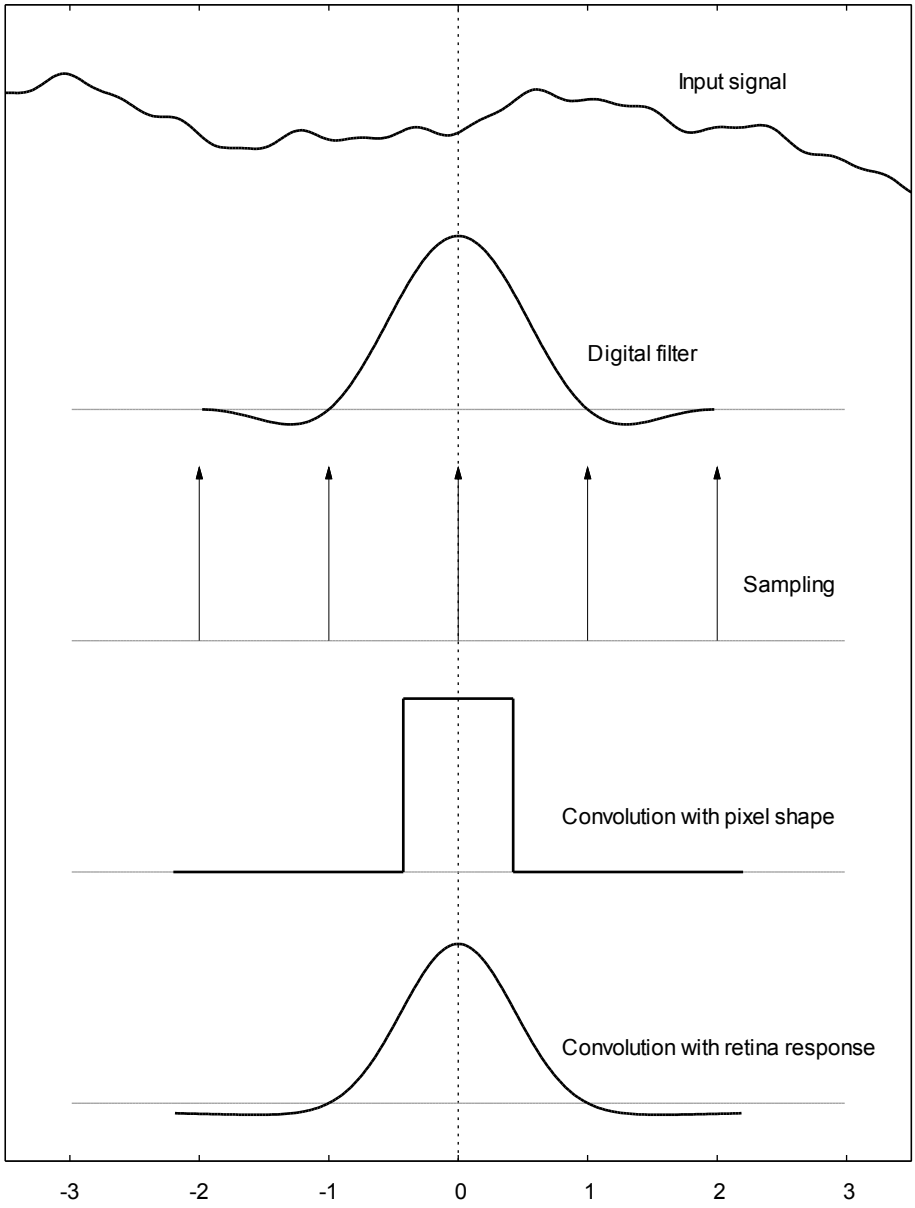


Figure 43: Scheme of the post-processing and observing process

For the image observed directly the sensation is

$$Sens_{direct}(x) = [Img * Eye](x) \tag{5.33}$$

The best filter would give the same result as direct observing. However this is not possible due to the loss of information by sampling. The filter can by only optimized. Unfortunately the optimal *Filter* with minimal error can not be generally expressed. The equation contains both convolution and multiplying, so minimization can not be solved in spatial nor in frequency domain. But the problem could be split into different sub-pixel positions *s*. The result needs to be expressed as a convolution for any case of s. Each case gives a different *Kernel*, so the complete operation is not convolution. But later we can minimize the error across all kernels.

$$x = n + s \mid n \in \mathbb{Z}, s \in \langle -\frac{1}{2}, \frac{1}{2} \rangle$$
$$Sens(n+s) = [Img * Kernel_s](n+s) \tag{5.34}$$

Now we are looking for the $Kernel_s$. One convolution and multiplying with the dirac impulses (1) can be expressed as a sum.

$$Sens(n+s) = \sum_{i \in \mathbb{Z}} [Pixel * Eye](n+s-i) \cdot [Img * Filter](i) \tag{5.35}$$

Let us define a notation for shifting the function parameter:

$$f_{(r)}(x) = f(x+r) \tag{5.36}$$

According to the convolution definition, shifting of the result argument is equal to shifting of one of the convolved function's argument. Here *i* and *(n+s)* are swapped.

$$Sens(n+s) = \sum_{i \in \mathbb{Z}} [Pixel * Eye](n+s-i) \cdot [Img * Filter_{(i-(n+s))}](n+s) \tag{5.37}$$

The convolution is a linear operation. Therefore the sum of convolutions, where one of the operand is constant, is a convolution with the sum of the other operands. This gives us the *Kernel*.

$$Sens(n+s) = [Img * Kernel'_{s,n}](n+s)$$
$$Kernel'_{s,n} = \sum_{i \in \mathbb{Z}} [Pixel * Eye](n+s-i) \cdot Filter_{(i-(n+s))} \mid n \in \mathbb{Z} \tag{5.38}$$

This *Kernel* needs an extra parameter *n*, but as *n* belongs to integers and in both occurrences is subtracted from *i* which is an argument of sum across all integers, *n* can by simply omitted.

$$Kernel'_{s,n} = Kernel_s = \sum_{i \in \mathbb{Z}} [Pixel * Eye](s-i) \cdot Filter_{(i-s)} \tag{5.39}$$

Error Minimization Method

As was mentioned above, the ideal filter would give the closest result to direct observing. It means that all variants of $Kernel_s$ would be same as *Eye*. The error of the sensed image can not be enumerated for general data. For this reason the error of the Kernel has to be used. It can be expressed as the absolute difference integral.

$$Error_s = \int_{x=-\infty}^{\infty} |Kernel_s(x) - Eye(x)| \, dx \tag{5.40}$$

Difference square is often used to express error, but square is related to the signal energy. In this case the sensation depends linearly on the image.

The space of all Filters has to be searched by brute-force algorithm. For this purpose the space has to be reduced reasonably. Following method was used:

- *Filter* is designed via Fourier transform. The highest harmonic frequency is not above the spatial frequency recognizable by the HVS. The amplitudes are not complex numbers. Filter should be symmetrical (even function), so only cosine harmonics are contained. This reduces the parameters to a relatively small amount of numbers.
- The parameter quantization step was selected as 1/1000. For the 8-bit displays the precision is sufficient. The *Eye* absolute values below 10^{-4} were ignored.
- Filter area (integral) should be 1. It gives that the zero harmonic value F_0 is inverse of the Filter size.
- The Filter value at both ends should be 0. It gives that $F_0-F_1+F_2-F_3+\ldots = 0$
- The spatial domain quantization was 11 samples per pixel. It is dense enough, so that the highest sampled frequency according to Nyquist is well beyond HVS recognition and odd number made some of the numeric computations easier.

The only value we do not know is the size of *Filter*. Let us submit a hypothesis that *Filter* of 2 pixels radius is sufficient. The error minimization process was run twice, for the *Filter* of radius 2 and 3 pixels.

Optimization Result

The results are compared in Figure 44. It is clear that the function is close to zero from the eccentricity about 2 pixels.

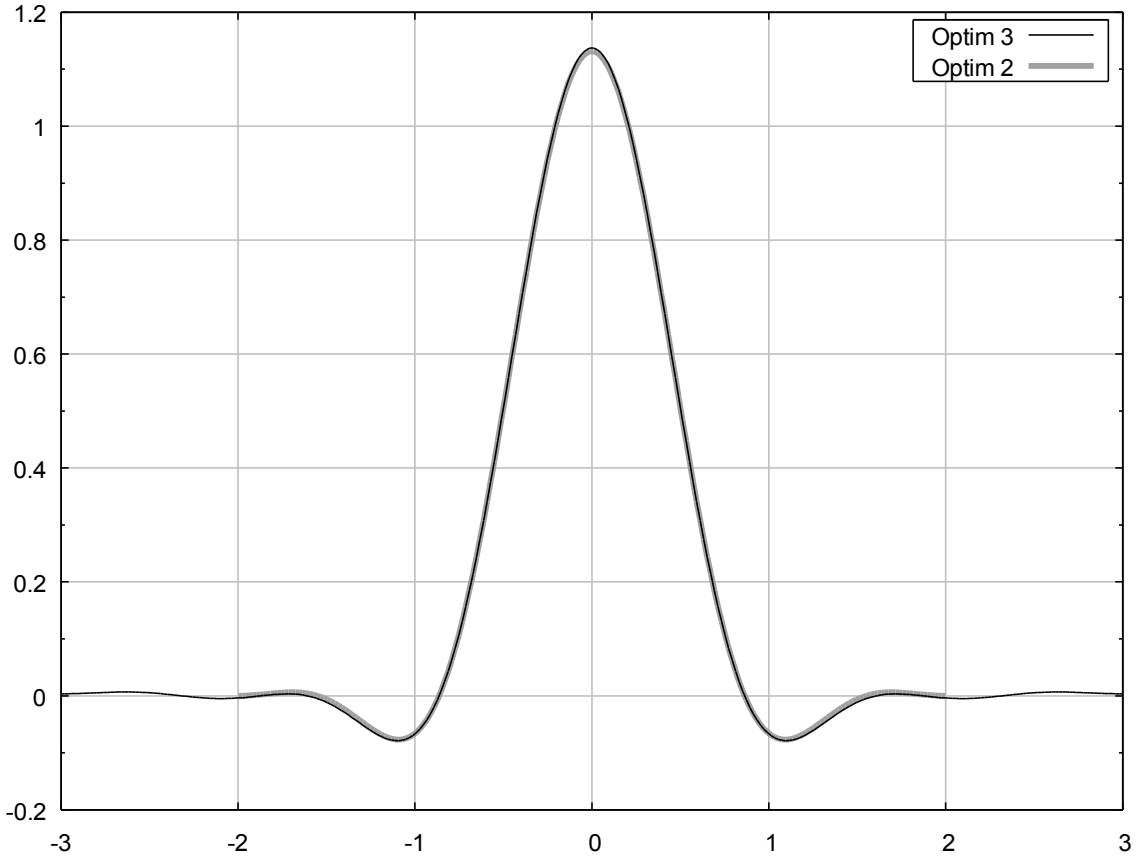


Figure 44: Optimization results for different filter sizes

The amplitudes of the filter with 2 pixels radius given by the optimization are listed below:

$$
\begin{aligned}
&Filter(x) = \\
&\quad 1/4 + \\
&\quad 0.504 \cdot \cos(1 \cdot \pi/2) + \\
&\quad 0.302 \cdot \cos(2 \cdot \pi/2) + \\
&\quad 0.048 \cdot \cos(3 \cdot \pi/2)
\end{aligned}
\tag{5.41}
$$

Nonlinearity Problem

The resampling method with the optimized filter was implemented and tested. The images were subjectively slightly sharper than with using traditional method, but the edges did not appear over-enhanced like with the using of hi-pass filter. But some barely visible artifacts were still apparent; very tiny, dark contours around some edges were detected. Note, that all images were processed as values of linear luminance with no gamma correction. It seems fitting the fact that the luminance is perceived logarithmically and so it might have been processed in the logarithmic scale. The truth is much more complex. Firstly, the filter integrates the luminance and luminance must be only integrated in linear domain, because of the physical nature of the light. And secondly, as we concluded [25] from data of Purpura and Kaplan. [2], the perception depends linearly on the local contrast, and logarithmically on the global illumination. The optimized filter is to correct the local contrast. But the problem is: in spite of the fact, that the filter should change only the local contrast, the least affected area of the display is whole pixel. The pixel size of the used display (0.270mm) is still about twice the sigma of the projected receptive field center Gaussian. So the question is: Could the pixel area take part in global illumination perception? Unfortunately we could find no answer in present results. The current knowledge about perceptual quality of digital images [24] does assume that the HVS nonlinearity is independently placed before the contrast perception and that different spatial frequency stimuli are perceived in the same nonlinear scale. This statement is now obsolete, taking into account Purpura and Kaplan's measurement. The source of the scale duality has not been named yet and till some more accurate description is found, we may continue only with a hypothesis.

We suggest to split the kernel into two parts. The convolution with the parts is calculated in the linear scale. The results are summed in the logarithmic scale. The size of each part should be large enough to interfere global illumination. Much smaller size than a pixel would also raise the integrating problem in nonlinear scale.

Results

The comparison of the error function is difficult, because no other resampling filter was optimized via our approach. If the frequency analysis of the whole system is measured (the system includes resampling filter, pixel shape and retina spatial response), the system passes approximately 5% higher frequency compared to the system with traditional Lánczos filter. This is not phenomenal improvement, but the number does not cover reduced artifacts and overall image quality. Due to mentioned insufficiency in methods for perceptual quality comparison, precise evaluation is currently not available.

5.7 Summary of Results

The performance of the scanning correction was improved with a new algorithm for fast geometry correction. The algorithm enables for implementation in a programmable hardware. Within the constraints of the approach, it preserves the image, as well as the more complex implementations, but with considerably higher performance. The performance of the dynamic range reduction was improved with new implementation of bilateral filter approximation. The filter results were compared to the exact bilateral filter implementation limited to the radius 5σ. The precision was measured on twenty-nine different images, each with the area approximately 0.7megapixels. The intensity sigma was set according to two state-of-the-art approaches [5, 6]. PSNR did not drop below

43dB for $\sigma i = 4$dB [6] or below 69dB for $\sigma i = 0.6$dB [5]. While the exact bilateral filter time dependency on the image area is almost exactly quadratic, in the approximation method the dependency is close to linear.

The perception was improved via new resampling optimization method. The method exploits a study of the display observation in many subjects. The proposed approach was used to measure correlation between optimal observation distance from the display device and the visual acuity. A standard monitor with the pixel spacing 0.270mm was used. The correlation was measured on 20 subjects. The relative acuity measured at the optimal observation distance of each subject was 3.98 cycles·mm^{-1}, standard deviation in logarithmic scale was 0.024 $\log_{10}$ cycles·mm^{-1}. The statistics were used to project the retina cell receptive field to the display plane. A novel resampling filter optimization was designed. The optimization respects spatial response of retina neural cells, measured relative acuity and shape of the display device pixel.

6 Conclusions

The reconstruction of an image via digital display is a complex problem. The result is still not ideal with commonly available devices and current methods. The aim of this work was to identify the weak points in the whole system and improve them. Some of the methods already provide transfer with error undetectable by human vision, but the processing is too slow for interactive view or real-time video processing. Several improvements were achieved and described in this work.

A new resampling technique for correcting geometry distorted images was presented. The design enables for hardware realization. The technique allows for very fast resampling and can be performed even real-time on high resolution pictures. A new approximation of bilateral filter was proposed. Bilateral filter is a core of several state-of-the-art methods for dynamic range reduction and it's computation is the critical part in the performance. The new approximation is computed in constant time per image pixel for theoretically any size of the image. The error did not exceed the theoretical threshold considered as an error detectable by human observer for any tested image.

The optimal observation distance from the display device was examined experimentally. A correlation between optimal distance and subject's visual acuity was measured statistically on a group of people. The measured correlation was used in a novel approach of image resampling for display devices. The method exploits present knowledge about human vision physiology, said experiments with observation distance and shape of the display pixel. Visual sensation is computed via visual system model and optimized for least difference from the original image. As far as we know, image processing optimization involving all these aspects was not attempted so far.

Several weak points of the image reconstruction framework were improved. The framework with proposed changes enhances both performance and perceived image quality.

The bilateral filter acceleration is fast, but enables for implementation using parallelism. Our intent is realization of high performance HDR processing using some modern parallel hardware architecture, e.g. NVIDIA® CUDA™. We expect that the implementation will be able to process high resolution pictures real-time. The resampling optimization asks for further investigation. The observation distance measurement ought to be done for higher density displays, which are becoming more available. The relation between the display density and optimal distance is to be approximated in order to design a general optimization of the resampling algorithm. Further investigation is also expected in the field of vision physiology. The source of the HVS dual luminance scale is now unclear. This aspect of human vision should be also reflected in the image processing, once the problem is enlightened. The discussed resampling method for a display with spatially split base colors [28], in our opinion, lefts a space for improvement. Base colors displacement is used, but the error is not enumerated and no alias effect or physiology behavior is considered. We intent to design an optimized resampling method for such displays.

7 References

[1] Croner, L.J., Kaplan, E.: Receptive Fields of P and M Ganglion Cells Across the Primate Retina, Vision Res. Vol. 35, No. I, pp. 7 24, 1995

[2] Purpura, K., Kaplan, E., Shapley, R.M.: Background Light and the Contrast Gain of Primate P and M Retinal Ganglion Cells, Proc. Natl. Acad. Sci., Vol. 85, pp. 4534-4537, Jun 1988

[3] Schwartz, S.H.: Visual Perception: a Clinical Orientation, 3rd ed., McGraw-Hill, ISBN: 0-07-141187-9, 2004

[5] Ledda, P., Santos, L.P., Chalmers, A.: A Local Model of Eye Adaptation for High Dynamic Range Images, Proc. of the International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction, AFRIGRAPH '04, pp. 151-160, ISBN: 1-58113-863-6, 2004

[6] Durand, F., Dorsey, J.: Fast Bilateral Filtering for the Display of High-Dynamic Range Images, Proc. of the Conference on Computer Graphics and Interactive Techniques, USA, pp. 257-266, ISSN: 0730-0301, 2002

[7] Seeman, M., Zemčík, P., Juránek, R., Herout, A.: Fast Bilateral Filter for HDR Imaging, J. Vis. Commun. Image R., Vol. 23, pp. 12-17, 2012

[8] Ferwerda, J.A.: Elements of Early Vision for Computer Graphics, IEEE Computer Graphics and Applications, pp. 22-33, ISSN: 0272-1716, Sep-Oct 2001

[9] Osterberg, G.: Topography of the layer of rods and cones in the human retina, Acta Ophthal Vol. 6, pp. 1-103, 1935

[10] Peichl, L., Wassle, H.: Size, Scatter and Coverage of Ganglion Cell Receptive Field Centres in the Cat Retina, J. Physiol. 291, pp. 117-141, 1979

[11] Hammond, P.: Cat Retinal Ganglion Cells: Size And Shape of Receptive Field Centres, J. Physiol. 242, pp. 99-118, 1974

[12] Dacey, D., Packer, O. S., Diller, L., Brainard, D., Peterson, B., Lee, B.: Center Surround Receptive Field Structure of Cone Bipolar Cells in Primate Retina, Vision Res. Vol. 40, Issue 14, pp. 1801-1811, Jun 2000

[13] Nelson, R., Kolb, H., Freed, M.: Off-Alpha and Off-Beta Ganglion Cells in Cat Retina I: Intracellular Electrophysiology and HRP Stains. J. Comp. Neurol. 329, pp. 68-84, 1993

[14] Kolb, H.: How the Retina Works, Am. Scientist, 91, pp. 28-35, Jan-Feb 2003

[15] Pridmore, R.W.: Bezold-Brucke Hue-Shift as Functions of Luminance Level, Luminance Ratio, Interstimulus Interval and Adapting White for Aperture and Object Colors, Vis. Research, Vol. 39, Issue 23, pp. 3873-3891, Nov 1999

[16] Lamming D.: Spatial Frequency Channels. Chapter 8. In: Cronly-Dillon, J., Vision and Visual Dysfunction, Vol 5., 1991

[17] Van Nes F.L., Bouman, M.A.: Spatial Modulation Transfer in the Human Eye, J. Opt. J. Optical Society of America, Vol. 57, Issue 3, pp. 401-406, 1967

[18] Anderson, S.J., Mullen, K.T., Hess, R.F.: Human Peripheral Spatial Resolution for Achromatic and Chromatic Stimuli: Limits Imposed by Optical and Retinal Factors, J. Physiol. 442, pp. 47-64, 1991

[19] Dowling, J. E., Boycott, B.B.: Organization of the Primate Retina: Electron Microscopy, Proc. R. Soc. 166, pp. 80-111, 1966

[20] Kolb, H.: Organization of the Outer Plexiform Layer of the Primate Retina: Electron Microscopy of Golgi-Impregnated Cells, Phil. Trans. R. Soc. 258, pp. 261-283, 1970

[21] Liang, J., Williams, D.R., Miller, D.T.: Supernormal Vision and High-Resolution Retinal Imaging Through Adaptive Optics, J. Opt. Soc. Am. A14, pp. 2884-2892, 1997

[22] Derrington, A.M., Krauskopf J., Lennie, P.: Chromatic Mechanisms in Lateral Geniculate Nucleus of Macaque, J. Physiol. 357, pp. 241-265, 1984

[23] Pridmore, R.W.: Extension of Dominant Wavelength Scale to the Full Hue Cycle and Evidence of Fundamental Color Symmetry. Color Res. Appl., 18, pp. 47–57, 1993

[24] Daly, S.: The Visible Differences Predictor: an Algorithm for the Assessment of Image

Fidelity, Digital Images and Human Vision, pp. 179-206, MIT Press 1993
[25] Seeman, M., Zemčík, P.: Vision Physiology Survey for Image Reproduction and Manipulation, submitted to Computers & Graphics
[26] Seeman, M., Zemčík, P.: Visual Acuity and Comfortable Distance from a Display, Poster Proc. of the 20th Int. Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision WSCG, 2012
[27] Seeman, M., Zemčík, P.: Improving Image Perception on Display Devices, submitted to J. Vis. Commun. Image R.
[28] Hill, W., Duggan, M., Keely, L.B., Hitchcock, G.C., Whitted, J.T.: Methods and Apparatus for Performing Image Rendering and Rasterization Operation, US Patent 6307566B1, 2001
[29] Yamada, E.: Some Structural Features of the Fovea Centralis in the Human Retina, Arch. Ophthal. 82, pp. 151-159, 1969
[30] Peichl, L., Wassle, H.: Size, Scatter and Coverage of Ganglion Cell Receptive Field Centres in the Cat Retina, J. Physiol. 291, pp. 117-141, 1979
[31] Seeman, M., Zemčík, P.: Histogram Smoothing for Bilateral Filter, Proc. GraVisMa, Plzen, Czech Rep., pp. 145–148, 2009
[32] Yoshizawa, S., Belyaev, A.G. Yokota, H.: Fast Gauss Bilateral Filtering, Computer Graphics Forum, 29, pp. 60–74., 2010
[33] Reinhard, E., Devlin, K.: Dynamic Range Reduction Inspired by Photoreceptor Physiology, IEEE Trans. on Visualization and Computer Graphics, Vol. 11, No. 1, pp. 13-24, 2005
[34] Zemčík, P., Přibyl, B., Herout, A., Seeman, M.: Accelerated Image Resampling for Geometry Correction, J. Real-Time Image Proc., Vol. 6, No. 3, ISSN 1861-8200, 2011
[35] Gouras, P.: Color Opponency From Fovea to Striate Cortex, Investigative Ophthalmology, Vol. 11, No. 6, Jun 1972
[36] Gallagher, A.C.: Detection of Linear and Cubic Interpolation in JPEG Compressed Images, Proc. of the 2nd Canadian Conference on Computer and Robot Vision, Vol. 171, pp. 65-72, 2005
[37] Zemčík, P., Herout, A., Seeman, M., Přibyl, B.: Způsob a Zařízení pro Digitální Korekci Obrazu, Czech Republic Patent 2010-650, 2010