Image Processing for Improved Perception and Interaction

Michal Seeman
Department of Computer Graphics and Multimedia
Faculty of Information Technology
Brno University of Technology
Božetěchova 2, 612 66, Brno, Czech Republic
seeman@fit.vutbr.cz

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.

Categories and Subject Descriptors
I.4 [Computing Methodologies]: Image Processing and Computer Vision

Keywords
visual perception, image processing, optimization, acceleration

1. Introduction
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 substantial change in the contrast causes noticeable change in the color perception [8].

2. Perception Optimizing on a Display
The scheme in Figure 1 describes the contributions to the image processing process presented in this work.

2.1 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.
Figure 2: 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.

The algorithm exploits separable resampling via FIR filter bank (see Figure 2). The set of filters is for selection of the subpixel displacement.

This approach enables for implementation using a pipeline with low consumption of resources in a programmable hardware. 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.

2.2 Dynamic Range Reduction Acceleration

The tone mapping operators for dynamic range reduction has been rapidly improved during last decade. One of the most complex physiologically influenced methods is [5]. This method (and many others, e.g. [3]) 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 [11] [6] [7] [12], in 2011 we designed an approximation method with very small error and fastest computation so far. The method is presented here.

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. 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.

Unlike in most other attempts to accelerate the filter, the image is split spatially in the presented approach:

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.

We proposed the method for histogram filtering. 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 filter results were compared to the exact bilateral filter implementation limited to the radius 5\( \sigma \). The precision was measured on twenty-nine different images, each with the area approximately 0.7megapixels. An example of the differential image is in Figure 4. The intensity sigma was set according to two state-of-the-art approaches [5] [3]. PSNR did not drop below 43dB for \( \sigma_i = 4dB \) [3] or below 69dB for \( \sigma_i = 0.6dB \) [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 (see Figure 3).

2.3 Resampling

It has to be said, that ideal resampling is not necessarily best for display devices. The reasons are:
Figure 4: (a) Tone-mapped input image, (d) differential image (Ledda’s σ, single EMA [5])

Figure 5: Perception of an image on a display: Image is filtered and sampled, rendered via the display pixels and processed by the HSV

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 low-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.

The resampling should respect shape of the display pixel, spatial response of the visual system and observation distance. These problems are discussed below.

2.4 Relation Between Visual Acuity and Optimal Observation Distance

Users tend to view the display from the so-called "comfortable distance". The question is how does the comfortable observation distance correspond to the visual acuity. The proposed approach was used to measure correlation between optimal observation distance from the display device and the visual acuity.

Users were to compare image post-processing methods. Tiny differences forced them to carefully choose optimal distance (see Figure 6). The distance was then measured by triangulation. Visual acuity was measured at the display surface, so the accommodation conditions were comparable (the optotypes detail os shown in Figure 7). A standard monitor with the pixel spacing 0.270mm was used. The correlation was measured on 20 subjects.

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

Although the acuity varies, it shows strong correlation with the preferred observing distance (Figure 8 and Figure 9). The results show that the relative spatial acuity in preferred distance has much smaller deviation than the angular acuity.

The statistics were used to project the retina cell receptive field to the display plane.

2.5 Resampling Filter Optimization

When the image is post-processed, rendered on a display and observed, the whole process can be described as shown in Figure 5 and Figure 10. The spatial response of the human visual system is not simple. It is formed by complex neural network within the retina. The retina contains photoreceptors and neural cells. The signal is processed by several tens of specialized cell types [4] [10], which forms the typical center-surround spatial response [2]. Measured characteristics of the primate visual system [1] [9] (see Figure 11) can be used to optimize the resampling.

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.
Figure 8: Correlation between angular acuity and preferred observing distance

Figure 9: Distribution of angular acuity across the age

Figure 10: Scheme of the post-processing and observing process

Figure 11: Small eccentricity P ganglion cell spatial response [1] recalculated from section to one-dimensional integral.

The process contains both convolution and multiplying, so minimization cannot be solved in spatial or 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 convolution kernel, so the complete operation is not convolution. But we can minimize the error across all kernels.

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 visual system spatial response 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 $F_0-F_1+F_2-F_3+... = 0$

- 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 error minimization process was run twice, for the filter of radius 2 and 3 pixels. The results are compared in Figure 12. It is clear that the function is close to zero from the eccentricity about 2 pixels. The amplitudes of the filter with 2 pixels radius given by the optimization are listed below:

$$\text{Filter}(x) = 1/4 + 0.504 \cdot \cos(1 \cdot \pi/2) + 0.302 \cdot \cos(2 \cdot \pi/2) + 0.048 \cdot \cos(3 \cdot \pi/2)$$
3. 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. The framework with proposed changes enhances both performance and perceived image quality.

References


Selected Papers by the Author