S·E·K·E·R: TONE REPRODUCTION BASED ON THE HUMAN VISUAL SYSTEM

ANSON LOPEZ, Oscar; GUTIERREZ PEREZ, Diego; SERON ARBELOA, Francisco Jose

Instituto de Investigación en Ingeniería de Aragón
Departamento de Informática e Ingeniería de Sistemas
Universidad de Zaragoza, España
Grupo de Informática Gráfica Avanzada
Correo electrónico: diegog@unizar.es

RESUMEN

Las técnicas existentes para generar imágenes sintéticas calculan la luminancia de las escenas. Los dispositivos que se usan para mostrar los resultados (monitor, video, papel, transparencias, etc.) generalmente no pueden reproducir las luminancias calculadas ya que se sitúan fuera del rango que pueden mostrar dichos dispositivos. Está claro que el éxito de las imágenes fotorealistas depende tanto del cálculo correcto de las luminancias como de su mapeo al dispositivo de salida para provocar en el observador la misma sensación. Esta conversión de las luminancias reales a las luminancias del monitor se llama reproducción de tono. Otro factor a tener en cuenta en la creación de imágenes sintéticas es el sistema visual humano (HVS): efectos como la saturación de los fotorreceptores, el deslumbramiento velante o la pérdida de color afecta a la manera en la que percibimos el mundo que nos rodea y no se puede calcular en la fase de rendering. En esta comunicación presentamos S·E·K·E·R, una aplicación que mapea luminancias reales a luminancias de pantalla considerando, además, las limitaciones del sistema visual humano y sus efectos en las imágenes finales.

Palabras clave: Reproducción de tono, sistema visual humano, realismo, imágenes de Alto Rango Dinámico, metodologías y técnicas.

ABSTRACT

The existing techniques to generate synthetic imagery calculate the luminances of the scenes. Devices used to display those results (monitor, video, paper, slide, etc) generally can not reproduce the computed luminances, since they usually fall outside of the display range. It is clear that the success of realistic images depends both on the correct calculation of the luminances and the correct mapping of those luminances to the display, in order to provoke in the observer the exact same response. This conversion from real luminances to display
luminances is called tone reproduction. Another factor to bear in mind in the creation of synthetic imagery is the mechanisms of the Human Visual System (HVS): effects such as bleaching of photoreceptors, veiling glare or color loss affect the way we perceive the world around us, and can not be calculated at the rendering stage. In this paper we present S·E·K·E·R, an application that correctly maps world luminances to display luminances while taking into account the limitations of the HVS and their effect on the final images.

Key words: Tone reproduction, human visual system, realism, High Dynamic Range imagery, methodologies and techniques.

1. Introduction

The objective of the synthetic fotorealistic imagery is to exactly capture the visual appearance of the modelled scenes. Physical based rendering methods allow to calculate with accuracy the energy distribution in the scene, and lots of works have been directed to these problems. However, this exact calculation does not guarantee that the visual appearance of the displayed image matches the real scene. This fact is because of two reasons: on one hand, the range of luminances in a real scene usually overcomes, in several magnitude orders, the usable range of the display device. Pictures in a newsletter have a maximum contrast of 30:1, CRT monitors have a standard range of 100:1, and only some high photographic quality printings have ranges of 1000:1. However, it is easy to found light contrasts of 100.000.000:1 in the real world. On the other hand, visualization conditions of the real scene and synthetic image rarely matches. In addition, studies in human visual mechanisms have not come to a definitive model [1][2].

The application of the human perception mechanisms to the synthetic imagery also allows a saving in calculation times. Understanding how the mind is going to get the scene, solutions can be then calculated with lesser accuracy in physical metrics terms, but knowing that a more physical exact solution (and more time consuming) will contribute nothing to the final image as it is perceived by the observer. This fact is very interesting in knwoledge areas, such as Virtual Reality, where images have to be calculated in real time. Some metrics based on perception to measure the quality of imagery can be founded in [3].

It is important to be highlighted that the main problem of the tone reproduction is the reduction of the contrasts in the image, mantaining its appearance. All the other effects, such as color sensitivity loss, visual acuity, chromatic adaptation and temporal responses directly depend on the selected solution to adapt the contrast. The diversity of methods suggest that it does not exist an only correct solution for the tone reproduction problem. Display devices only can suggest the sensations and contrasts that they can not reproduce, and they do it based on various elements. It is possible, therefore, that several combinations of these elements give a valid solution.
2. High Dynamic Range Imagery

Most of the graphic hardware works with a 24 bit (per pixel) RBG color space, setting 8 bit (1 byte, a range from 0 to 255) for each channel (red, green and blue). The advantage of this representation is that we do not need any tone reproduction technique to obtain reasonable results in any kind of CRT device, since every value within the range [0, 255] match with a luminance level of the monitor (also depending on a characteristic curve, called gamma curve) [4]. On the other hand, there is one disadvantage: all the colors that fall out of the RGB gammut can not be displayed. That is because of the fact that the dynamic range of a standard monitor is about 100:1 (approximately 2 magnitude orders). Dynamic range is the ratio of maximum and minimum luminance. On the other hand, a human bserver can perceive a dynamic range up to 4 or 5 log units while the adaptation process and up to 9 log units when he is totally adapted. Finally, luminance values in the real world can reach ranges of 10 log units. This means that RGB color space only cover a little range of the luminance range that can be perceived in the real world.

A simple solution that is used is to work with floating values, allowing the use of values above the typical range of [0, 255]. That minimizes the loss of information if we use real luminance values. But on the other, this raw representation has an important disadvantage with the size it needs to be stored: 32 bits per floating value, a pixel size of 96 bits, in contrast to the 24 bits per pixel in the RGB color space. Fortunately there are many representations based on the human perception, which allow to work with 32 bit floating pixels, even smaller [5][6][7].

There are two general methods to generate High Dynamic Range (HDR) images: physical based rendering and taking photographs (with a common camera) with several exposure levels of a real scene [8]. It is expected that in a future, camera manufacturers give support to the HDRI techniques and principles. Therefore, computer rendering is nowadays the more straight solution. If a rendering system is based on physical phenomenons [9], it will calculate the real radiance values in each pixel of the image. These spectral values will have to be converted to displayable values, work done by the tone reproduction operator.

3. Human Visual System

Our visual system is able of distinguishing between a huge light range. We only have to think about the ratio of a dark night and a sunny day (up to 10,000,000:1). The light goes into the eye and stimulate the two types of photoreceptors in the retina; they are called cones and rods. Rods are very sensitive to the light intensity and they are more active in scotopic light levels (from $10^{-6}$ to $10^{-2}$ \text{cd/m}^2). Cones are less sensitive and give us chromatic vision in photopic light levels (from $10^{2}$ to $10^{8}$ \text{cd/m}^2), were they are totally active. That is why we can not distinguish color in darkness, where we only can see in gray levels. Both systems are are simultaneously active in the range from $10^{2}$ to $10 \text{cd/m}^2$. This range is called mesopic range and is the more unknown among the three and also the more important one.
Figure 1: Real world range of luminances and associated visual parameters.

After stimulating photoreceptors, the signal generated by photoreceptors goes through brain cells over the brain, where the image is formed. While the eye can receive up to 14 log units of light, optic nerve can only transmit 1.5 log units. That means that there is any kind of adaptation in our visual system. Once we are adapted to a luminance adaptation level we then can perceive data around this adaptation level [10]. Through this adaptation process, our visual system controls the effects of the illumination change in the visual response to provide the appropriate sensitivity in a wide range of environmental illumination levels.

Physiological mechanisms are the base for the psychophysical description of the adaptation process. The action of these mechanisms is reflected in several visual effects: glares, bleaching, color loss, visual acuity, temporal sensitivity.

4. S·E·K·E·R

S·E·K·E·R, the application that we present in this paper, is a tone reproductor that maps images from real luminances to display luminances and simulate several limitations of the human visual system. We have follow two criterions to do a reliable tone reproduction: on one hand, it conserves the visibility reproduction. That means that an object can be seen in the image if and only if it can be seen the real scene. On the other hand, the image provoke a subjective sensation in the observer. It reproduces the sensation of bright, acuity, contrast and color.

Our application is based on the work of [11]. S·E·K·E·R generates the adaptation luminances histogram of the HDR image and modifies it to discover clusters of adaptation levels. Afterwards, human visual limitation models are applied to simulate the effects of the human visual system. Finally the image is mapped to match the human contrast sensivity.

Adaptation histogram

The generation of the histogram starts filtering down the original image, getting a smaller image in which each pixel matches with an area of 1° in the original image (a potential fixation point or foveal area). This image is also called foveal image. The histogram is generated from the luminances of this foveal image. Once generated the histogram, S·E·K·E·R modifies it with a human contrast sensitivity function [10] to
simulate the human response to contrast. With simpler methods (histogram equalization, etc.), contrast could be exaggerated in highly populated zones of the histogram. Basically, the human contrast sensitivity function reduces contrast in dark zones of the image as our visual system does.

From the samples of the adaptation histogram a cumulative distribution function (CDF) is generated. CDF will be used at the last stage to map the original luminance image.

![Figure 2: General workflow for the S·E·K·E·R application.](image)

**Human visual limitation pipe**

As it can be seen in Figure 2, there are two different processes in S·E·K·E·R: the upper horizontal one is the generation of the foveal image, the adaptation histogram and the final mapping function. In the bottom of the figure appears the human visual limitation pipe. This pipe modifies the original image to add three different human limitations: visual acuity loss, veiling glare and color sensitivity loss.

Bright sources in a scene reduce the perception of the global contrast in a scene, because of the dispersion of light in the lens darken the fovea (the main area of vision in the eye). S·E·K·E·R uses the methods of [12] and [13] that calculate effective adaptation luminance from the position of bright sources and illuminance in the scene. Therefore, on one hand, a veiling image is calculated from the foveal image. Afterwards we add this veiling mask to the original foveal image to obtain a new foveal veiled image, in which each pixel is the effective adaptation luminance. At this moment S·E·K·E·R can regenerate the new adaptation histogram to obtain the correct mapping function. On the other hand, the veiling image is used, by means of extrapolation, to add the veiling glare to the original image.

To simulate the color loss in dark environments, S·E·K·E·R uses the method described in [10]. This limitation is also applied in the original image and follows a simple operation: for each pixel in the original image, if its luminance falls in the
scotopic range, rods system is more sensitive than cone system and we have color loss, so pixel is converted to a grey scale; if it falls in the photopic range, cones prevail and our visual system has full color vision, so pixel is kept as it is. If the pixel luminance falls within the mesopic range, we use a linear interpolation based on scotopic and photopic luminances.

As well as losing the skill of seeing the contrast and the color, human eye can not distinguish details in dim scenes. To simulate this visual limitation S-E-K-E-R uses data from [14] and a local blurring function that blur those zones of the image with dim luminances. This blurring function is implemented as a filter of variable resolution and works with a pyramid image and parametric interpolation, based on the mip mapping works of Williams [15].

5. Results

The image that we use as an example of what S-E-K-E-R can do has been obtained from [16]. Figure 3 shows a false color map of the luminance levels in the image in log units. Red colors show a high level of light and blue colors, on the contrary, show the dark zones of the image. The scene is basically the interior of a cathedral. The main light sources in the image are the glazed vault and three glass windows (all in red and yellow).

Figure 3: False color map of scene luminances
Figure 4 shows the generated adaptation histogram. Horizontal axis shows the adaptation luminances in the scene in log units. Vertical axis shows the number of adaptation samples in the foveal image. The adaptation luminance range is from 0.2 cd/m² to 25000 cd/m².

Figure 4: Adaptation histogram

Figure 5 show the scene mapped with a simple linear mapping function. From top to bottom and from left to right, S-E-K-E-R-simulates larger and larger exposure times. So, the first image is under-exposed and the last one is over-exposed. This would be the simplest method to map a HDR image.

Figure 5: Linear tone reproduction with several exposure times
Finally, Figure 6 shows the difference between a simple tone mapped image (left) and the same scene simulating all the human visual limitations (right): loss of color and visual acuity in dark areas and veiling glare in bright zones. It also simulates the real human contrast perception.

Figure 6: Comparison of a basic tone reproduction (left) with a full simulation of human visual limitations (right).

6. Conclusions

While a great work has been done in the development and improvement of the physical based rendering algorithms, tone reproduction and spectral rendering techniques do not have the same level of evolution and they still have important opened questions to be interpreted and investigated.

Although the advance in the displaying technologies affords us better and better devices, we still depend on tone reproduction operators to get a desired perceptual interpretation of the synthetic images. Even so, all the display devices in the market go on having very narrow dynamic ranges and it is expected that the future of tone reproduction depends on the development of devices with broader dynamic ranges. Other important opened question is the storage of the HDR images. Although there are existing techniques that manage storing real light values, the ratio of size and accuracy goes on being inversely proportional.
Finally we have the studies and knowledge of the HVS, which go on being quite limited, especially in its neural stage. Moreover, the accurate modelling of their features in the tone reproduction methods is complex and very time consuming. The lack of understandable and reliable metrics for the image analysis also limits the study.

7. Future work

The method to calculate foveal samples for the fixation points of the observer can be extended. It will depend on set up elements by means of an interactive system or a preplanned animation. It could be applied the gaze theory of the eye (probable attention directions) to improve the initial adaptation histogram. Additional modifications could be done in sensitivity threshold, glare and visual acuity models to simulate the effects of the age in an observer.

Other factor that has not been treated is the temporal behavior of the adaptation process. It could be implemented a new module that simulates the HVS when it has to adapt to light or darkness, based on [10] and [17] models. Other improvement refers to changes in the color sensitivity in the mesopic range. S-E-K-E-R implement this change as a linear interpolation between both response functions for the scotopic and photopic range [10], but a better approximation should be done. The influence of the luminous environment also should be regarded in the adaptation process of the observer. Finally, the implemented sensitivity color loss model does not consider the absolute perception of the color, strongly affected by the global adaptation and the color of the light sources.

The tone reproduction method could be extended to other application areas. S-E-K-E-R can be incorporated in global illumination calculations to make them more efficient.

The work that will be carried out in the future will be the application of the tone reproduction techniques to the immersive environments visualization. Precisely, the objective of tone reproduction is to move closer synthetic perceptual sensations to the real ones in a real environment. But monitors can not get this total immersion for the observer in the synthetic scene. That is why we want to use these tone reproduction techniques in an immersive environment, as it is the CAVE, designed by the GIGA [18] [19]. In this way, the observer, apart from feeling immerse in a virtual scene, can have a more real perception of the environment.

References


