Chapter 7

Depth Estimation for HDR Images

S. Manikandan
Electronics and Radar Development Establishment, Defense Research and Development Organization, India

ABSTRACT

In this chapter, depth estimation for stereo pair of High Dynamic Range (HDR) images is proposed. The proposed algorithm consists of two major techniques namely conversion of HDR images to Low Dynamic Range (LDR) images or Standard Dynamic Range (SDR) images and estimating the depth from the converted LDR / SDR stereo images. Local based tone mapping technique is used for the conversion of the HDR images to SDR images. And the depth estimation is done based on the corner features of the stereo pair images and block matching algorithm. Computationally much less expensive cost functions Mean Square Error (MSE) or Mean Absolute Difference (MAD) can be used for block matching algorithms. The proposed algorithm is explained with illustrations and results.

INTRODUCTION

We normally see the world in three dimensions. This is because each eye looks a slightly different view of a scene and the brain converts the information into a 3D image. Stereo pair images contain depth information due to the parallax inherent in the images. Thus it is possible to extract depth information from this stereo image pair. But in image processing, computer graphics and in photography, High Dynamic Range Imaging (HDRI) is a set of techniques that allows a greater dynamic range of luminance between lightest and darkest areas of an image than standard digital imaging techniques or photographic methods. This wider dynamic range allows HDR images to more accurately represent the wide range of intensity levels found in real scenes, ranging from direct sunlight to faint starlight. The two main sources of HDR imagery are computer renderings and merging of multiple photographs, which in turn are known as Low Dynamic Range (LDR) /
Standard Dynamic Range (SDR) photographs. Tone mapping techniques, which reduce overall contrast to facilitate display of HDR images on devices with lower dynamic range, can be applied to produce images with preserved or exaggerated local contrast for artistic effect.

The natural world presents our visual system with a wide range of colors and intensities. A starlit night has an average luminance level of around $10^{-3}$ candelas/m$^2$, and daylight scenes are close to $10^5$ cd/m$^2$.

Humans can see detail in regions that vary by $1:10^4$ at any given adaptation level, over which the eye gets swamped by stray light (i.e., disability glare) and details are lost. Modern camera lenses, even with their clean-room construction and coated optics, cannot rival human vision when it comes to low flare and absence of multiple paths (“sun dogs”) in harsh lighting environments. Even if they could, conventional negative film cannot capture much more range than this, and most digital image formats do not even come close. With the possible exception of cinema, there has been little push for achieving greater dynamic range in the image capture stage, because common displays and viewing environments limit the range of what can be presented to about two orders of magnitude between minimum and maximum luminance. A well-designed CRT monitor may do slightly better than this in a darkened room, but the maximum display luminance is only around 100 cd/m$^2$, which does not begin to approach daylight levels. A high-quality xenon film projector may get a few times brighter than this, but they are still two orders of magnitude away from the optimal light level for human acuity and color perception.

The human eye has two different types of photoreceptors. Cones are responsible for sharp chromatic vision in luminous conditions, or the photopic range. Rods provide less precise vision but are extremely sensitive to light and allow us to see in dark conditions, or the scotopic range. Both rods and cones are active in moderately luminous conditions, known as the mesopic range.

Light adaptation, or simply adaptation, is the (fast) recovery of visual sensitivity after an increase or a small decrease in light intensity. Otherwise, the limited range of neurons results in response compression for (relatively) high luminances. This is why everything appears white during the dazzling observed when leaving a tunnel. To cope with this and to always make the best use of the small dynamic range of neurons (typically 1 to 40, sensitivity is controlled through multiplicative (gaincontrol) and subtractive mechanisms. It is reasonable for the steady state, however since subtractive and multiplicative mechanisms have different time-constants and different effects, they should be differentiated for a more accurate simulation of adaptation. Dark adaptation is the (slow) recovery of sensitivity after a dramatic reduction in light. It can take up to tens of minutes. The classic example of this is the adaptation one experiences on a sunny day upon entering a theater for a matinee. Initially, everything inside appears too dark, and visual acuity is, at best, poor.

Note that all of these mechanisms are local, i.e. they occur independently for single receptors or for small “pools” of receptors. This is motivated by efficiency considerations, but also because considering local adaptation states is extremely challenging. The interaction between local adaptation and eye-gaze movements is very complex and is left as a subject of future research. However, choosing a single adaptation level (or “average” light intensity) is not trivial.

Chromatic adaptation allows us to perceive objects as having a constant color, even if they are observed under illuminants with varying hues. This type of adaptation involves mechanisms in the retina as well as higher level mechanisms. It can be reasonably modeled as a different gain-control for each cone type, which is called Von Kries adaptation. Chromatic adaptation can be complete or partial and takes approximately one minute to completely occur. One of the fascinating aspects of this type of adaptation is that it is not driven by the average color of the stimulus, but