https://doi.org/10.2352/ISSN.2470-1173.2017.13.IPAS-086

Video frame synthesizing method for HDR video capturing system with four image sensors

Takayuki Yamashita; Ehime University; Matsuyama city, Ehime Prefecture; NHK(Japan Broadcasting Corp.), Tokyo/Japan
Yoshihiro Fujita; Ehime University; Matsuyama city, Ehime Prefecture/Japan

Abstract

High Dynamic Range (HDR) imaging has recently been applied to video systems, including the next-generation ultrahigh definition television (UHDTV) format. This format requires a camera with a dynamic range of over 15 f-stops and an S/N ratio that is the same as that of HDTV systems. Current UHDTV cameras cannot satisfy these conditions, as their small pixel size decreases the full-well capacity of UHDTV camera image sensors in comparison with that of HDTV sensors.

We propose a four-chip capturing method combining three-chip and single-chip systems. A prism divides incident light into two rays with intensities in the ratio m:1. Most of the incident light is directed to the three-chip capturing block; the remainder is directed to a single-chip capturing block, avoiding saturation in high-exposure videos. High quality HDR video can then be obtained by synthesizing the high-quality image obtained from the three-chip system with the low saturation image from the single-chip.

Herein, we detail this image synthesis method, discuss the smooth matching method between spectrum characteristics of the two systems, and consider the modulation transfer function (MTF) response differences between the three- and single-chip capturing systems by means of analyzing using human visual models.

Introduction

In recent times, High Dynamic Range (HDR) imaging has been applied not only to still images, but also to video sequences. A significant difference in the HDR system of still images is that the dynamic range on display side is also expanded to both end of lower and brighter.

Some standards for next generation television systems include the HDR video and ultrahigh definition (UHDTV) formats. In the case of HDR systems, a dynamic range of over 15 f-stops is required for the camera; in addition, a better SN ratio than that of a Standard Dynamic Range system is required. However, current UHDTV cameras cannot satisfy the aforementioned conditions, because the full-well capacity of UHDTV cameras is lower than that of HDTV cameras.

To address these issues, A number of methods have been proposed. One is to expand the dynamic range of an image sensor itself. This approach inclides a method that utilizes MOSFET's sub-threshold characteristics [1-3]. Another method uses the non-destructive read-out function of an active pixel sensor [4]. we proposed another approach of using a four-chip capturing method that combines the ordinary three-chip capturing system used in broadcasting applications with a single-chip capturing system [5]. The incident light is divided into two light rays by a prism with intensities in the ratio m:1. Then, the low light that is refracted by the prism is directed into a single-chip capturing block to avoid saturation in the case of high exposure video shooting. By synthesizing the high quality picture from the three-chip capturing system with the low saturation picture obtained from the single-chip capturing system, an HDR video with high quality can be obtained.

In this paper, we discuss this picture synthesizing method in detail. First, we discuss the smooth matching method of spectrum characteristics between the three-chip system that obtains colors using a tricolor separation prism and the single-chip system that obtains colors through the on-chip color filter. Second, we consider the absorption of MTF difference between the two capturing systems at the transition point between the three-chip capturing and single-chip capturing systems.

We analyze these issues using human visual models and show appropriate synthesizing method and required parameters and values for synthesizing.

Basic Concept of HDR video capturing system with four image sensors[5]

Optical system

Color cameras, especially those for broadcasting for which high picture quality is required, have relied primarily on three-chip color imaging since the advent of Phillips type prisms. This method, which uses these prisms, offers several advantages. Cameras based on this imaging are highly sensitive owing to small incident light losses. Resolution is also high as each of the three colors (Red, Green, Blue) is assigned its own image sensor. Color reproducibility is good owing to optimal spectral characteristics; suited for television standards. These advantages make this three-chip coloring imaging by far the best method available.

When obtaining two images with different exposures, the use of three-chip prisms is desirable if high picture quality is required. Time division can be employed to obtain these two images, but moving images produced by this method have choppy quality, making them unsuitable for television. Ideal picture quality can be obtained by using two pairs of three- chip optical systems to obtain two images, one with low exposure and the other with high exposure, but this approach makes the optical system unduly complicated.

A solution is the proposed method that combines three-chip color imaging and a single-chip color imaging. As the incident light travels through the prism block (Figure 1), it is reflected by the first prism surface (the reflected amount is expressed by 1/(m+1), where m is the exposure ratio) before reaching the single-chip color image sensor. The remaining amount, m/(m+1), is further reflected by the second and third surfaces as it travels through the prism block, getting divided into Red, Blue, and Green in the process before reaching the respective image sensors.

If m is 1 or greater, the side of single-chip color imaging is for low exposure imaging and the side of three-chip color imaging is for high exposure imaging, meaning that the single-chip side handles the high intensity light while the three-chip side handles the low intensity light. This relationship will be reversed if m is set smaller than 1. As mentioned above, images taken by the three image sensors have high picture quality, while the picture quality of images taken by the single-chip color imaging is not so high. Generally, images with low intensity light often take up a major part of the screen in the case of assured scenes; it is necessary in such a case to set m greater than 1. Sugawara had devised a four-chip image sensor camera [6] designed to provide high resolution. Though the minimum F value of this camera is slightly greater, it has about the same optical path length and color reproducibility as the three-chip color imaging system.

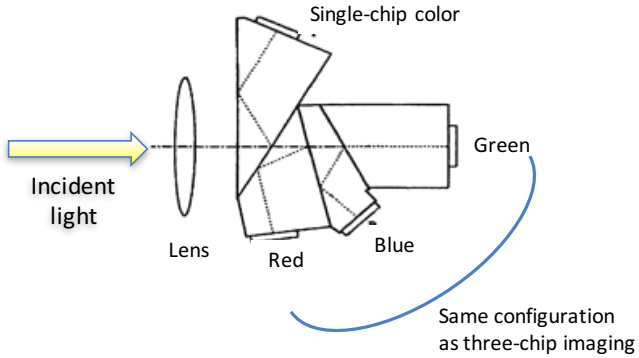

figure 2. Optical block in this system

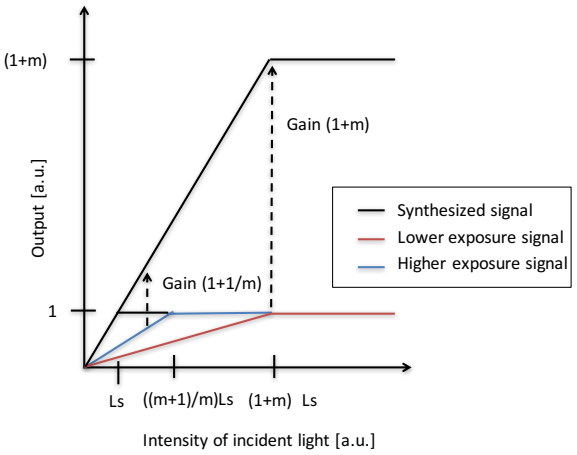

Figure 1. Transfer function characteristics

Method of Image Synthesizing

Let us consider a scene in which low intensity light takes up the main part and high intensity light is assigned to the minor portion. The principle of image synthesis is to replace the saturated highlight portion of high quality images obtained by the three-chip color imaging with high-intensity light images taken by the single-chip color imaging. To synthesize two images, the switching point is set in the area of high intensity light near the reference white level.

Here, let us suppose that each pixel of a single-chip color imager has the same maximum charge capacity as each imager for the tri-color imaging does. It is also supposed that the spectrum characteristic of color filter array is as same as that of the three-chip prism.

Figure 2 shows input-output characteristics, where L_{in} is the level of the incident light, L_H is the output level on the high-exposure side, and L_L is the level on the low-exposure side. If the input-output conversion gain is 1, L_H and L_L are expressed by the following.

$$L_H = \frac{m}{m+1} L_{in} \tag{1}$$

$$L_L = \frac{1}{m+1} L_{in} \tag{2}$$

With L_S, being the incident light level at which L_o saturates, the amount of incident light on the high-exposure side at which L_H saturates is given by $(m+1)/mL_S$ considering that the amount of incident light becomes $m/(m+1)$ after separation as shown in equation (1). In the same manner, the amount of incident light at which L_L saturates is given by $(m+1)\ L_S$. To synthesize two images, L_H and L_L are each multiplied by amplification factors "m+l/m" and "m+1" before making a switch from high-exposure images to low-exposure images at the saturation point of L_H. In this way, the incident light will remain linear until $(m+1)\ L_S$ meaning that the dynamic range is (m+1) times. The expansion rate of the dynamic range increases with the ratio of exposure m.

There are two HDR standards in video industries, Perceptual Quantization [7] (PQ) and Hybrid Log-Gamma [8] (HLG). Both of standards are included in Recommendation for international program exchange format in ITU-R [9].

The feature of PQ is that the digital codes are related with absolute value of display luminance. On the other hand, the feature of HLG is that the digital codes are related with relative levels of image sensor output. Design of HLG succeeded to that of the current Standard dynamic range（SDR）system.

Opto-Electronic transfer function(OETF) of HLG in Rec. ITU-R BT.2100 is represented as follows:

$$E' = \begin{cases} \sqrt{E}/2, & 0 \le E < 1 \\ a \cdot \ln(E-b) + c, & 1 < E \end{cases} \tag{3}$$

(a = 0.17883277, b = 0.28466892, c = 0.55991073)

where E is the signal for each color component proportional to scene linear light and scaled be camera exposure, normalized to the range [0:12]. E' is the resulting non-linear signal in the range [0:1]. From this equation, it is found that the maximum dynamic range is 1200 %.

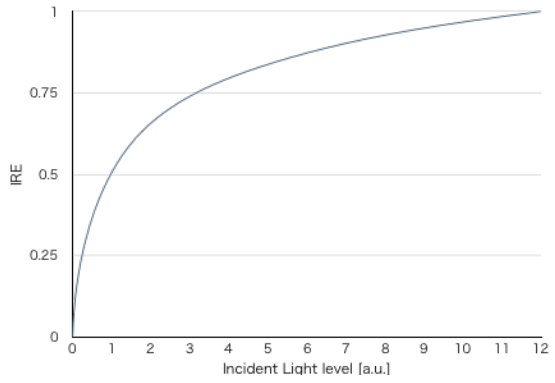

Figure 3. HLG OETF

Figure 3 shows the chart of this OETF. Considering our proposed system is applied to HLG, m equals 11.

Analysis of SN Ratio

When value of m sets bigger dynamic range of the camera becomes higher but its SN ratio deteriorates, because the low-exposure image is amplified. To understand this problem further, we examined dark noise (a portion not dependent on incident light) and photon shot noise.

In the following discussion, N_{dk} und N_{st} represent, respectively, dark noise and photon shot noise at one imager. When the incident light is divided into m:l, dark noise N_{ldk} and photon shot noise N_{lst} on the low exposure side are given by:

$$N_{ldk} = N_{dk} \tag{4}$$

$$N_{lst} = \frac{1}{\sqrt{m+1}} N_{st} \tag{5}$$

When the incident light is l/(m+1), the photon shot noise deteriorates in proportion to its square root and we obtain equation (5). The incident light being amplified by (m+1) when the images are combined, equations (4) and (5) will then become:

$$N_{ldk} = (m+1)N_{dk} \tag{6}$$

$$N_{lst} = \sqrt{m+1}N_{st} \tag{7}$$

We can see that the dark noise increases in proportion to m+1, and the photon shot noise in proportion to $\sqrt{m+1}$. By combining these two noises, we obtain the total amount of noise as follows:

$$\begin{aligned} N_{Total} &= \sqrt{{N'_{ldk}}^2 + {N'_{lst}}^2} \\ &= \sqrt{\{(m+1)N_{dk}\}^2 + {(m+1)N_{st}}^2} \end{aligned} \tag{8}$$

SN ratio is given by:

$$(SN\ ratio) = S \Big/ \sqrt{\{(m+1)N_{dk}\}^2 + {(m+1)N_{st}}^2} \tag{9}$$

When this system use in SDR video system, as the white level is customarily suppressed by knee processing for broadcasting cameras, SN ratio is better than this. It is necessary to take these into consideration in determining the separation ratio m.

Let consider that this system use in HDR video system. For simplicity, assuming that the image sensor has a 12-bit output and the noise is less than 1 LSB, the SN ratio is 72 dB. When m = 11 derived earlier, gain increase by 12 times is achieved, so S / N degradation of -22.3 dB is obtained, which is equivalent to 50 dB. Figure 4[10] shows the measured values of the image noise detection source and the group of different dashed and solid lines showing the relationship between the brightness of the display device, but nearly the same result is derived. From this chart, it can be seen that if SN ratio is 50 dB or more, Noises are not detected in an image of 100 rlx. BT.2100[9] specifies that a mastering monitor with a peak of 1000 cd/m^2 should be used.

Therefore, when calculating the incident light by calculating (without total system gamma) using inverse OETF as follows, a 100rlx of the luminance level on a monitor corresponds to 50 % of a level of incident light, 100 % of a level of the incident light as the switching has sufficient margin.

$$\begin{aligned} E &= \mathrm{OETF}^{-1}[E'] \\ &= \begin{cases} 4E'^2, & 0 \le E' \le \frac{1}{2} \\ \exp\left(\frac{E'-c}{a}\right) + b, & \frac{1}{2} < E' \end{cases} \end{aligned} \tag{10}$$

(The values of parameters a, b, and c are as defined for the OETF)

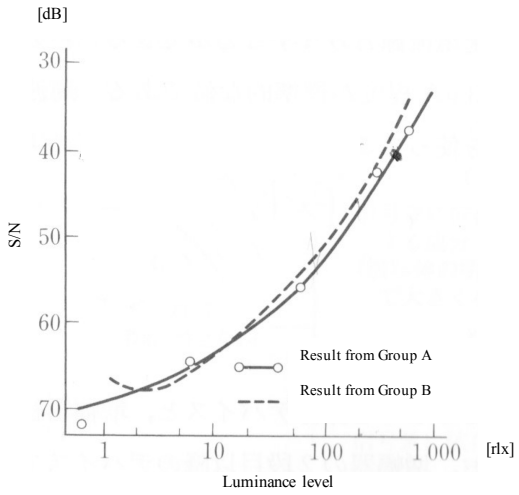

Figure 4. Relation between detection limit on S/N and Luminance level on monitor[]

Other issues to consider when synthesizing

When synthesizing video images, it is necessary to absorb the difference between the three methods because it is going to synthesize the three-chip imaging and the single-chip imaging. We will examine these methods here.

Difference of spectral responses

The three-chip imaging method is a color separation by a prism, and the other single-chip color imaging is coloring by a color filter using a dye on a chip. Regarding each coloring method, Reference [11] specifies the ideal characteristics, but since these spectral characteristics are different, color correction which is the same as general color matching method between cameras is necessary. Applying color correction by the 3D look up table to either the single-chip or the three-chip coloring output can absorb the difference.

Difference of MTFs

For the single-chip color imaging method, demosaicing for interpolating the information of each pixel is required from the Bayer pixel arrangement. Depending on the demosaicing method, the response of MTF generally tends to decrease as spatial frequency closes to the Nyquist frequency. In this system, since the low exposure side has only the low-pass characteristics, the characteristics on the high exposure side contribute to the scene image with a high contrast, so that the frequency characteristics are improved in the synthesized image.

Simulation

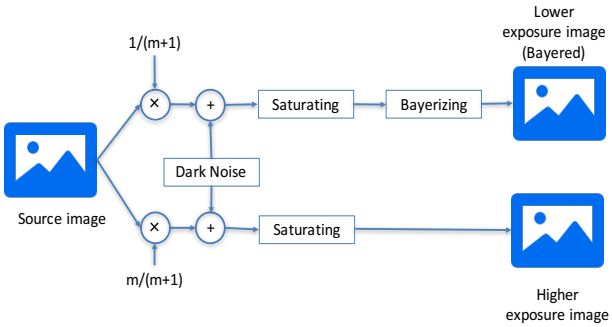

Figure 5. Acquisition model for simulation

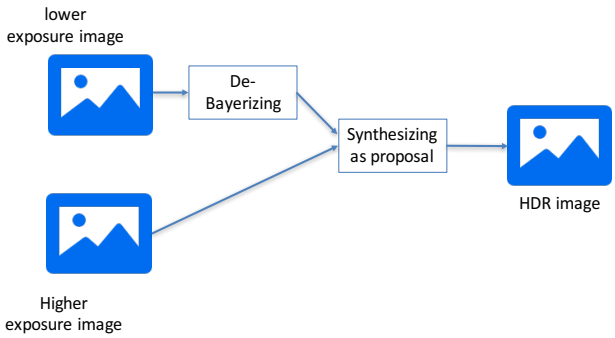

Figure 6. Synthesizing Model

We modeled and simulated the proposed system.

Figure 5 is a model of the imaging system simulating the optical system. Noise is added, and since the lower exposure side corresponds to a single plate, Bayer treatment is added.

On the other hand, a model for synthesis is shown in Figure 6. Gain is added to each of the lower exposure side image and the Higher exposure side image subjected to the de-bayer processing. For the switching point between the Low side and the high exposure side, the synthesized image is output as the level at which the higher exposure side saturates.

Figure 7. Source HDR image ("Dani_belgium_oC65", Image courtesy Dani Lischinski)

For this simulation, Radiance .hdr format image was used.(Figure 7) The synthesized image after signal processing is shown in Figure 8. Also, images on low exposure side and high exposure side during synthesis are also shown in Figures 9 and 10. It can be seen from these images that the output image is converted to the high dynamic range.

In addition, regarding the low noise property which is a feature of this system, it is evaluated by mean square error with the original image. This is shown in Table 1.

Figure 8. Synthesized HDR image

Figure 10. image on low exposure side (Gained)

Figure 9. image on high exposure side

Table 1 Image quality evaluation

Image	Root Mean Square Error to original image
Synthesized Image	21.2997
Image on Lower Exposure	41.9740
Image on Higher Exposure	799.5911

This table shows that the RMSE of the synthesized image is smaller than the image on the low exposure side and the high exposure side, and is close to the original image despite the addition of the Gaussian noise.

Conclusion

We proposed a new HDR acquisition system that combines three-chip color imaging and a single-chip color imaging and video synthesizing method. We analyzed the relationship between the separation ratio of the incident light and SN ratio. The proper value was derived from HDR-TV standard and we confirmed it based on human vision system. We have also simulated this system conducted this is effective in obtaining high-quality pictures with a high dynamic range.

References

[1] G. Chamberlain, et al., "A novel wide dynamic range silicon photodetector and linear imaging array," *IEEE Journal of Solid-State Circuits*, sc-19. No. 1, pp. 41-48, 1984.

[2] Spyros Kavadias, et al., "On-chip offset calibrated logarithmic response image sensor," *1999 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, pp. 68-71, 1999.

[3] M. Loose, et al., "Self-calibrating logarithmic CMOS image sensor with single chip camera functionality," *1999 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, pp. 191-194, 1999.

[4] Shimamoto, et al., "Dynamic Range Expansion using a CMD Imager," *The Journal of the Institute of Image Information and Television Engineers*, 54, No. 12, pp. 1781-1787, 2000.

[5] T. Yamashita, et. al., "Wide-dynamic-range camera using a novel optical beam splitting system," *Proc. SPIE 4669, Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications III*, 82, 2002

[6] Sugawara, et al., "Four-Chip CCD camera for HDTV," *SPIE PROCEEDINGS*, 2173, pp. 122-129, 1994.

[7] SMPTE ST 2084, "High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays," Society of Motion Picture & Television Engineers, 2016.

[8] ARIB STD-B67, "ESSENTIAL PARAMETER VALUES FOR THE EXTENDED IMAGE DYNAMIC RANGE TELEVISION (EIDRTV) SYSTEM FOR PROGRAMME PRODUCTION," Association of Radio Industries and Businesses, 2015.

[9] Recommendation ITU-R BT.2100, "Image parameter values for high dynamic range television for use in production and international programme exchange," International Telecommunication Union, 2016.

[10] Y. Nishida, et. al., "Noise in CCD image sensor and consideration for future system,", *ITEJ technical report*, vol. 9, no. 30, pp. 1-6, 1985. (in Japanese)

[11] ARIB TR-B37, "Interconnection for UHDTV Camera and Lens," Association of Radio Industries and Businesses, 2016.

Author Biography

Takayuki Yamashita is currently a doctoral student in Ehime University in Japan and a senior manager in Engineering Department at NHK (Japan Broadcasting Corp.) He is working on the research, development and standardization of ultrahigh-definition television systems. His research fields include the development of camera systems and the high-bandwidth digital signal processing. He joined NHK in 1995 and has been engaged in research of HDTV camera systems since 1999. He is a member of SMPTE, ATSC and the Institute of Image Information and Television Engineers of Japan (ITE).

Yoshihiro Fujita is currently a professor at the Dept. of Electrical and Electronics and Computer Science at Ehime University. He earned this position in 2011. From 1976 to 2011, he had been with Japan Broadcasting Corporation (NHK), where he conducted research on advanced imaging systems for HDTV and ultra-high definition TV (UHDTV). He received his B.E. and Ph.D. degrees in 1976 and 1998, respectively, from the University of Tokyo, and he is a fellow of IEEE.